

**Aus dem  
Department für Frauengesundheit Tübingen  
Universitäts-Frauenklinik**

**CTG-Beurteilung mit Methoden der Künstlichen Intelligenz**

**Inaugural-Dissertation  
zur Erlangung des Doktorgrades  
der Medizin**

**der Medizinischen Fakultät  
der Eberhard Karls Universität  
zu Tübingen**

**vorgelegt von  
Wolny, Agata Karolina, geb.Kleciek**

**2024**

Dekan: Professor Dr. B. Pichler

1. Berichterstatter: Professor Dr. H. Abele

2. Berichterstatter: Professor Dr. T. Küstner

Tag der Disputation: 29.04.2024

Diese Arbeit widme ich meinen Eltern, Anna und Zdzisław  
und meinem Ehemann, Adrian

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis</b>	<b>4</b>
<b>Abkürzungsverzeichnis</b>	<b>5</b>
<b>1 Einleitung</b>	<b>7</b>
1.1 CTG	7
1.1.1 Einführung und Vorgeschichte	7
1.1.2 Einsatzbereiche und Kritikpunkte	9
1.1.3 Automatisierte Auswertungssysteme	10
1.2 Künstliche Intelligenz	11
1.2.1 Das Konzept	11
1.2.2 CTG-Auswertung mittels KI: Literaturübersicht	13
1.3 Fragestellung und Ziel	14
<b>2 Materialien und Methoden</b>	<b>16</b>
2.1 Kardiotokographie - Prinzip der Methode	16
2.2 Auswertung des CTG nach den FIGO-Kriterien	18
2.3 Studiendesign und ethische Aspekte	19
2.4 Patientenkollektiv	20
2.4.1 Ein- und Ausschlusskriterien	20
2.4.2 Fallzahl und Qualitätssicherung	20
2.4.3 Maternale Charakteristika und Risikoprofil	22
2.4.4 Fetale Merkmale	23
2.4.5 Intrapartale Daten	24
2.5 Statistische Auswertung	26
2.5.1 Deskriptive Statistik und statistische Inferenz	26
2.6 Methoden der künstlichen Intelligenz	27
2.7 Aufbewahrung des Datenmaterials	27
<b>3 Ergebnisse</b>	<b>28</b>
3.1 Charakteristika der Studienpopulation	28
3.1.1 Subpopulation der Schwangeren	28
3.1.2 Fetales Kollektiv	29
3.1.3 Das Ereignis der Geburt	30
3.2 Patientendaten und deren Zusammenhang mit dem arteriellen Nabelschnur-pH-Wert	33
3.2.1 Gesamtpopulation	33
3.2.2 Subpopulation der spontanen Entbindungen	36
3.3 Datenanalyse mit Algorithmen der künstlichen Intelligenz	38
3.3.1 Kardiotokogramme und Patientendata der Gesamtpopulation	38
3.3.2 Blackbox Effekt	40

3.3.3 Subpopulation der sekundären Kaiserschnitt-Geburten	42
3.3.4 Kardiotokogramme und Patientendaten: Subpopulation der spontanen Geburten	43
3.3.5 Leistungsfähigkeit bei erhöhtem Anteil azidotischer Fällen	44
3.3.6 Re-Implementierung eines externen Modells auf eigenen Daten	45
<b>4 Diskussion</b>	<b>46</b>
<b>5 Zusammenfassung</b>	<b>51</b>
<b>Abbildungsverzeichnis</b>	<b>53</b>
<b>Tabellenverzeichnis</b>	<b>54</b>
<b>Literaturverzeichnis</b>	<b>55</b>
<b>Erklärung zum Eigenanteil der Dissertationsschrift</b>	<b>62</b>
<b>Danksagung</b>	<b>63</b>

## **Abkürzungsverzeichnis**

AFI - Amniotic Fluid Index

AI - eng. Artificial Intelligence

AUC - Area under the curve

BMI - Body Mass Index

CTG - Kardiotokographie, Kardiotokogramm

DGGG - Deutsche Gesellschaft für Gynäkologie und Geburtshilfe

DL - Deep Learning

ICSI - Intrazytoplasmatische Spermieninjektion

IUFT - Intrauteriner Fruchttod

KI - Künstliche Intelligenz

ML - Machine Learning

NS-pH-Wert - Nabelschnur-pH-Wert

Pc - Perzentile

PI - Pulsatility Index

SD - Single Depot

SpM - Schläge pro Minute

SG - spontane Geburt

SSW - Schwangerschaftswoche

STAN - ST-Strecken-Analyse

# 1 Einleitung

## 1.1 CTG

### 1.1.1 Einführung und Vorgeschichte

Die Kardiotokographie gilt weltweit als Standardverfahren zur intrapartalen Beurteilung des fetalen Wohlbefindens in der Schwangerschaft (präpartal) und unter der Geburt (intrapartal). Gegenüber der intermittierenden Auskultation der fetalen Herztöne mit beispielsweise dem Pinard oder Dopton erlaubt das CTG eine kontinuierliche Herztonregistrierung, die eine objektivierbare Dokumentation der Untersuchungsbefunde erlaubt. Bei der CTG-Registrierung handelt es sich um die simultane graphische Darstellung der Herzaktion des Fetus, der uterinen Kontraktionen und der Herzaktion der Mutter. Ziel der CTG-Untersuchung ist das rechtzeitige Erkennen von Herzfrequenzmustern, die auf eine drohende fetale Hypoxie und/oder Azidose hinweisen.

Die ersten Berichte über das Abhören der fetalen Herztöne mittels Stethoskop stammen aus dem 19. Jahrhundert. Im Jahr 1895 entwickelte der französische Arzt Adolphe Pinard das bis heute gebräuchliche Modell des geburtshilflichen Stethoskops, das umgangssprachlich sogenannte "Pinard-Rohr" (Dunn 2006). Die Überwachung mittels Pinard'schem Stethoskop benötigt die ständige Anwesenheit einer Fachperson und erfolgt dadurch überwiegend intermittierend, findet aber bis heute - insbesondere in der außerklinischen Geburtshilfe in Deutschland - Anwendung.

Ein erheblicher und nicht unumstrittener Impuls in der Geburtshilfe erfolgte durch die Erfindung des CTG-Gerätes durch den deutschen Gynäkologen Konrad Hammacher im Jahr 1968 (Bender et al. 2009). Der Kardiotokograph ermöglichte eine kontinuierliche Überwachung der Herztöne des Fetus, wobei sich das geburtshilfliche Fachpersonal (Hebammen, Ärztinnen und Ärzte) während der fetalen Herztonregistrierung auch auf andere Aspekte ihrer Arbeit konzentrieren konnte. Vor allem vor dem Hintergrund, dass die Geburtshilfe in vielen Ländern (mangels Ressourcen) keine reale 1:1-Betreuung von Gebärenden durch eine Hebamme ermöglichen kann.

Aus diesem Grund und wegen der guten Dokumentierbarkeit und der Möglichkeit der

simultanen Abbildung der fetalen Herzaktion mit Wehentätigkeit hat das CTG in der klinischen Geburtshilfe in den letzten 60 Jahren die intermittierende fetale Herztonauskultation ersetzt, obwohl in Studien nicht geklärt ist, zu welchem Zeitpunkt im Verlauf einer Geburt und in welchem Ausmaß eine CTG-Schreibung sinnvoll ist.

Ab den siebziger Jahren des 20. Jahrhunderts beobachtete man eine deutliche Senkung der Perinatalsterblichkeit in den industrialisierten Ländern. In Deutschland sank die Rate auf ein Siebtel, von 43 Fällen pro 1000 im Jahr 1955 auf unter 6 Fälle pro 1000 im Jahr 2020 (BIB, 2022). Viele Faktoren trugen zu dieser Verbesserung bei: unter anderem die Einführung der Asepsis, die Entwicklung der operativen Entbindungsmodi und die Verbreitung der neonatalen medizinischen Versorgung. Die Einführung der kontinuierlichen CTG-Überwachung wurde anfänglich als einer dieser Faktoren betrachtet. Eine Studie von K. Goeschen aus Hannover zeigte eine Halbierung der fetalen Mortalität in der Gruppe der mittels CTG überwachten Schwangeren, im Vergleich zu den Patientinnen, die ohne CTG-Überwachung betreut wurden (Goeschen 1991).

Im Laufe der Zeit kamen mehr Studien und neue Folgerungen hinzu und die Einschränkungen der Methode wurden vermehrt diskutiert. Die Sensitivität der Kardiotokographie beträgt 99 %, die Spezifität ist jedoch mit 77 % deutlich niedriger (Hösli et al. 2017). Das grundlegende Problem der Methode ist, dass eine normale Herztonaufzeichnung mit hoher Wahrscheinlichkeit auf ein Wohlbefinden des Kindes in utero hinweist, bei Normabweichungen der Herztonaufzeichnung die Prognose für den Fetus in manchen Situationen ungewiss ist. Die dadurch entstehenden falsch positiven Fälle führen zu unnötigen geburtshilflichen Interventionen im Sinne eines Kaiserschnitts oder einer vaginal-operativen Geburt.

Die Anwendung der Kardiotokographie scheint insbesondere in der Gruppe der Risikoschwangerschaften Vorteile zu haben. Allerdings sind die Kriterien für eine Risikoschwangerschaft weit gefasst, so dass mit der alleinigen Feststellung einer Risikoschwangerschaft nicht unmittelbar ein Vorteil der CTG-Aufzeichnung ableitbar ist.

Die oben genannten Kritikpunkte und vor allem eine hohe Subjektivität bei der CTG-Interpretation führten schon vor einigen Jahren zu den Ansätzen, die Auswertung des Kardiotokogrammes maschinell zu unterstützen. Es wurden Systeme entwickelt, die die Analyse der geburtshilflichen Fachpersonen verbessern

sollten (Oxford-CTG, STAN-CTG). Der Nutzen solcher Systeme bei der Betreuung des Gesamtkollektivs aller Schwangeren und Gebärenden ist jedoch unklar.

Die Frage nach einem Auswertungssystem, das im Normalkollektiv einsetzbar und nicht-invasiv anwendbar ist, lenkte die Aufmerksamkeit auf den sich dynamisch entwickelnden Bereich der Algorithmen der künstlichen Intelligenz.

### **1.1.2 Einsatzbereiche und Kritikpunkte**

Die antepartale Routine-CTG im Niedrigrisiko-Kollektiv wurde in den letzten Jahren zunehmend kritisch betrachtet. Die englische Leitlinie sprach sich gegen diese Indikation aus (Ayres-de-Campos et al. 2015). In Deutschland werden neue Leitlinien zu dem Thema im Jahr 2023 erwartet.

Eine der häufigsten Indikationen für die CTG-Anwendung ist die subpartale Situation. Die Studienlage bleibt umstritten. Ein pathologisches CTG-Muster wurde mit einem Risiko einer Zerebralparese (Nelson et al. 1996) und neonataler Enzephalopathie (Spencer et al. 1997) assoziiert. Eine Metaanalyse von Vintzileos belegt eine Reduktion der perinatalen Mortalität um über 50% (Vintzileos et al. 1995). Gleichzeitig weisen die Daten allerdings auf eine Erhöhung der Rate operativer Geburten um den Faktor 2.5 hin. Laut der letzten Cochrane-Metaanalyse senkt die kontinuierliche CTG-Überwachung im Vergleich zu der intermittierenden Herztonauskultation die Inzidenz von neonatalen Krampfanfällen. Die perinatale neonatale Mortalität, die Geburtsazidose und das Risiko einer Zerebralparese bleibt unverändert (Alfirevic et al. 2017). In Deutschland ist es seit Dezember 2020 zugelassen, bei unauffälligem Geburtsverlauf als Monitoring die intermittierende Herztonauskultation einzusetzen. Bei Auffälligkeiten wird eine kontinuierliche CTG-Überwachung empfohlen. Der meistdiskutierte Anwendungsbereich der CTG-Überwachung sind Niedrig-Risiko-Schwangerschaften. Was die Niedrigrisiko-Patientinnen im intrapartalen Setting betrifft, schränken die neuen DGGG-Leitlinien aus dem Jahr 2020 die Indikationen für eine Kardiotokographie in der Aufnahmesituation und in der beiden Phasen der Geburt im Vergleich zu der vorherigen Empfehlung ein (Abou-Dakn et al. 2020).

Anders ist die Lage im Hochrisiko-Kollektiv. In der Gruppe der Risikoschwangeren und bei Niedrigrisiko-Patientinnen mit intrapartalen Komplikationen wird die

kontinuierliche CTG-Anwendung unter der Geburt empfohlen. Obwohl die Daten für die Senkung der neonatalen bzw. perinatalen Mortalität und Spezifität bezüglich der Azidose in intrapartalem Setting den Ergebnissen im Niedrigrisiko-Kollektiv ähneln, sind die Empfehlungen im deutschsprachigen, englischsprachigen und internationalen Raum in diesem Patientinnen-Kollektiv übereinstimmend (Abou-Dakn et al.2020, NICE 2017, Ayres-de-Campos et al. 2015).

Gut dokumentiert und geprüft ist die CTG-Anwendung in antepartaler Situation bei hypotrophen Feten bzw. dem Vorliegen von Zeichen einer Plazentainsuffizienz. Eine der wichtigsten Studien in der Geburtshilfe, die multizentrische, randomisierte TRUFFLE-Studie hat erwiesen, dass die Überwachung mittels automatisiertem CTG (Oxford-CTG) und/oder Doppler-Sonographie die beste Kombination zur Optimierung der Entbindungszeitpunktes bei Feten mit intrauteriner Wachstumsretardierung ist (Bilardo et al.2017).

Was die medizinische Forschung zum Einsatz der antenatalen und perinatalen CTG-Registrierung betrifft muss angemerkt werden, dass viele Studienergebnisse alt sind und eine Wiederholung dieser Studien aufgrund ethisch-rechtlicher Aspekte (insbesondere im Kollektiv von Schwangeren) ethisch nicht vertretbar sind.

Bemerkenswert an der Entwicklung der Kardiotokographie ist die Tatsache, dass die Technik der Methode über die Jahre überwiegend unverändert geblieben ist. Veränderungen gab es dagegen hinsichtlich der Indikationen, der Auswertung und der Suche nach einer maschinellen Auswertungsmethode, die die menschliche Interpretation unterstützen könnte.

### **1.1.3 Automatisierte Auswertungssysteme**

Zwei wichtige Systeme, die durch eine computerisierte Analyse die Interpretation der Kardiotokographie unterstützen, sind das Oxford-CTG und das STAN-CTG. Das Oxford-CTG basiert auf den Dawes-Redman-Kriterien, die in den 80er Jahren von den beiden Oxforder Professoren Geoffrey Dawes und Chris Redman formuliert wurden. Einer der wichtigsten Parameter dieser Kriterien ist die Kurzzeitvariabilität. Es handelt sich um eine computerisierte Analyse der Abweichungen in Schlag-zu-Schlag-Variabilität, gemessen in Mikrosekunden (Pardey et al. 2002). Die Methode eliminiert auch den Messfehler, der bei subjektiver menschlicher Interpretation in der Standardanalyse entsteht. Die ersten Studien berichteten über

eine Korrelation zwischen Kurzzeitvariabilität und fetaler Azidose/Hypoxie in antepartaler Situation (Anceschi et al.2003). Die aktuelle Indikation wurde auf die Gruppe der Feten mit intrauteriner Wachstumsretardierung in der antepartalen Periode begrenzt.

Das STAN-System wurde das erste Mal Anfang des 21. Jahrhunderts in Schweden vorgestellt. Zur kontinuierlichen kardiotokographischen Messung wird hier eine ST-Strecke-Beurteilung aus dem fetalen Elektrokardiogramm angeschlossen. Die Anwendung erfordert eine zusätzliche fetale Kopfschwartenelektrode, was mit dem Risiko einer fetalen Infektion verbunden ist. Die Auswertung der ST-Strecke-Veränderungen erfolgt automatisch. Die Zielgruppe sind Schwangere unter der Geburt, Kontraindikationen sind u.a. intakte Fruchtblase sowie Frühgeburtlichkeit. Die Methode scheint im Vergleich zur alleinigen Kardiotokographie die Anzahl der operativen Entbindungsmodi zu senken, die Ergebnisse bezüglich der Vorhersage fetaler Azidose sind aber widersprüchlich (Becker et al. 2012). Die Möglichkeit der Anwendung nur nach dem Blasensprung, widersprüchliche Berichte bezüglich der Korrelation mit fetaler Hypoxie und der notwendige Gebrauch der Kopfschwartenelektrode waren Faktoren die Verbreitung des Systems in der geburtshilflichen Praxis limitierten.

In den letzten 25 Jahren wurden mehrere Systeme zur automatisierten CTG-Auswertung entwickelt, die auf den FIGO-Kriterien basieren. Nur eines davon, das englische INFANT-System, wurde in einer randomisierten kontrollierten Studie untersucht. Eine Verbesserung des kindlichen Outcome konnte nicht nachgewiesen werden, weshalb das System in der Routineanwendung nicht etabliert wurde (Brockehurst et al. 2017).

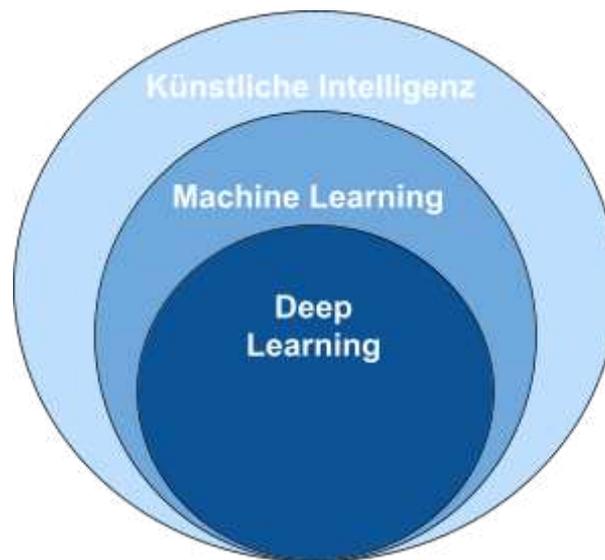
## **1.2 Künstliche Intelligenz**

### **1.2.1 Das Konzept**

Künstliche Intelligenz (KI, auch AI für engl. Artificial Intelligence) ist ein Teilgebiet der Informatik, welches sich mit der Automatisierung intelligenten Verhaltens und dem maschinellen Lernen befasst. Als Beginn dieses Wissenschaftsbereiches bezeichnet man das Jahr 1950, als der britische Mathematiker Alan Turing sich mit dem Konzept maschineller Intelligenz auseinandersetzte und den sogenannte "Turing-Test"

entwickelte. In dem Test sollte ein Mensch entscheiden, ob er sich mit einer Person oder einem Computer unterhält. Die Aufgabe war nicht lösbar, was bewies, dass ein Computer eine menschnenne Intelligenz besitzen kann (Copeland 2022).

Die zwei wichtigsten Gebiete der KI sind maschinelles Lernen (Machine Learning) und als Teilbereich davon tiefgehendes Lernen (Deep Learning). Die Beziehung der Begriffe zueinander kann wie in der untenstehenden Abbildung 1 dargestellt werden.



**Abbildung 1: Graphische Darstellung des Konzeptes der künstlichen Intelligenz**

Machine Learning (ML) ist ein Bereich der KI, in dem die Daten anhand von Algorithmen analysiert werden. Die KI lernt aus diesen Daten und wendet das Gelernte an. Ein gutes Beispiel sind Online-Anzeigen der Firma Google. Anhand von unseren Eingaben in die Suchmaschine werden unsere Interessen von dem Algorithmus erkannt, gelernt und danach für die gezielte Präsentation von Anzeigen verwendet.

Deep Learning (DL) ist hingegen eine Teilmenge von Machine Learning. DL strukturiert Algorithmen in Schichten, die ähnlich den biologischen neuronalen Netzen sind. Diese Algorithmen lernen und treffen intelligente Entscheidungen selbst. Ein interessantes Beispiel geben Forschungsarbeiten zur KI beim traditionellen chinesischen Spiel Go. DL erlaubte die Entwicklung eines Algorithmus, der der KI die Fähigkeit verlieh, sich ständig weiterzuentwickeln (zu lernen), bis sie

mit den besten Go-Spielern der Welt auf Augenhöhe spielen konnte.

Die Hauptunterschiede zwischen den beiden Gebieten sind eine geringere menschliche Interferenz und die mit den sich vergrößernden Datenbanken steigende Leistung von DL im Vergleich zu ML.

Zu den häufigsten Anwendungsgebieten künstlicher Intelligenz gehören Computer Vision, Natural Language Processing, Speech Recognition und Robotik. Die Verwendung von KI im medizinischen Bereich ist ein junges, dynamisches Wissenschaftsgebiet, welches die Verbesserung des ärztlichen Entscheidungsprozesses zum Ziel hat.

### **1.2.2 CTG-Auswertung mittels KI: Literaturübersicht**

Die wichtigsten Studien zum Thema CTG-Auswertung mit Methoden der künstlichen Intelligenz (Machine Learning und Deep Learning), die für die Entwicklung des hier beschriebenen Projektes essentiell waren, werden in diesem Kapitel besprochen. Die Übersicht konzentriert sich auf Studien mit Algorithmen, die ohne ärztliche Mitbeurteilung entwickelt worden sind. Die Übereinstimmung fachlicher Beurteilungen von Kardiotokogrammen ist mit 48 % lediglich zufriedenstellend, was den Problemlösungsprozess negativ beeinflussen kann (Hruban et al. 2015).

Der Zusammenhang zwischen dem intrapartalen pathologischen CTG und der fetalen Hypoxie/Azidose ist, bei bisher ungenügender Spezifität für normabweichene Fälle, ein potentieller Ansatzpunkt für den Aufbau konstant wachsender Datenbanken und wegen der Messmethode des CTG und den damit verbundenen zahlreichen Variablen für eine Analyse mit KI-Methoden prädisponiert. Entsprechende Vorarbeiten von Warrick et al. aus dem Jahr 2010 (ML-Methode) erlauben das Erkennen pathologischer Geburtsverläufe in über der Hälfte der Fälle (Warrick et al. 2010). Die Pathologie wurde als Basenüberschuss  $\geq 12$  mmol/l, Tod oder neurologische Symptome für ischämische Enzephalopathie bezeichnet. Diese Arbeit ist ein Beispiel aus dem Machine-Learning-Bereich, in dem eine Auswertung in kleinen Datenbanken (hier 264 Fälle) realisiert wurde.

Im Hinblick auf die aktuellen Entwicklungen der Analysemethoden der KI, den sogenannten "neuronalen Netzen" (DL-Bereich), muss vor allem die Arbeit von Petrozziello aus dem Jahr 2019 mit einer Datenbank erwähnt werden, welche über

35000 Fälle beinhaltet. Für die Risikoabschätzung der postpartalen fetalen pH-Werte  $< 7.05$  erreichte das multimodale faltungs-neuronale Netzwerk einen AUC-Wert von 0.77 (Petrozziello et al.2019).

In einer der neuesten Arbeiten von Ogasawara aus dem Jahr 2021 mit einer Fallzahl von 5406 Geburten erzielten die Autoren mit Hilfe eines tiefen neuronalen Netzwerks nennenswerte Ergebnisse. Die AUC betrug 0.73 für die Vorhersage arterieller pH-Werte unter 7.2 (Ogasawara et al. 2021) . Zu betonen ist hier jedoch ein höherer Cut-off-Wert im Vergleich zu der Arbeit von Petrozziello.

Eine Übersicht der aktuellen Arbeiten über Herausforderungen bei der Entwicklung von KI-Algorithmen zur CTG-Auswertung wurde in einem Review-Artikel von M. O'Sullivan präsentiert (O'Sullivan et al. 2021a). Der gleiche Autor prüfte in einer anderen Studie aus dem Jahr 2021 die Nutzbarkeit der klinischen Patientendaten in der Vorhersage des fetalen pH-Wertes und zeigt eine Verbesserung der Ergebnisse nach Einschluss einiger dieser Parameter (O'Sullivan et al. 2021b). Die Arbeit basiert jedoch nicht auf einer eigenen Datenbank, sondern auf der öffentlich verfügbaren Datenbank der Universitätsklinik in Brno (CTU-CHB).

### **1.3 Fragestellung und Ziel**

Die Daten bezüglich der Vorhersage einer fetalen Hypoxie und/oder Azidose anhand eines pathologischen CTG-Musters bleiben kontrovers. Vor allem in intrapartaler Situation, in denen das CTG in der klinischen Geburtshilfe neben der Überwachung und Anamnese eines der wichtigsten Entscheidungsinstrumente bleibt, ist die Datenlage bezüglich der Effektivität der Methode nicht zufriedenstellend.

Ziel dieser Studie ist es daher, über die Anwendung der neuesten Methoden der künstlichen Intelligenz (neuronale Netzwerktechnik) einen CTG-Auswertungsalgorithmus zu suchen, der zur Entdeckung einer intrapartal drohenden fetalen Hypoxie/Asphyxie im Niedrig- und Hochrisiko-Kollektiv genutzt werden kann.

Das geplante Design dieser Arbeit setzt im Unterschied zu den vorherigen Studien die Integration der klinischen Daten voraus. Dieses Vorgehen wurde bereits in den Werken der Arbeitsgruppe aus Oxford diskutiert (Petrozziello et al.2019) und in der

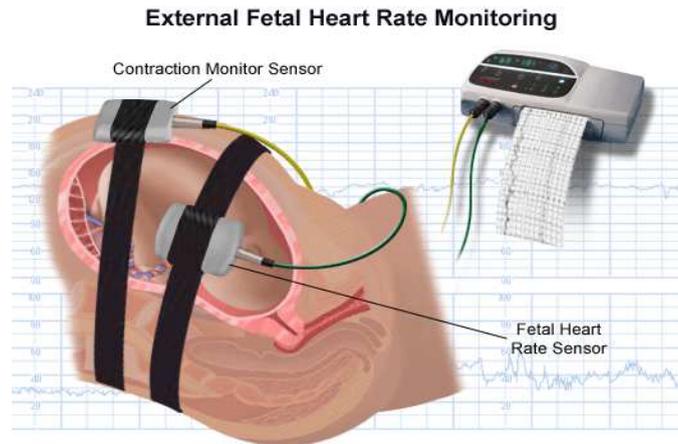
neuen Arbeit von O'Sullivan mit einem eingeschränkten klinischen Datensatz geprüft (O'Sullivan et al. 2021b). In unserer Studie wird erforscht, welche von den häufig klinisch genutzten Daten eine Bedeutung in der Vorhersage der fetalen Azidose und/oder Hypoxie haben könnten.

Das endgültige Ziel ist die Erstellung eines Instruments, das in der alltäglichen Praxis angewandt werden kann und die Einschätzung des zu erwartenden Nabelschnur-pH-Wertes bei Geburt in vivo ermöglicht.

## 2 Materialien und Methoden

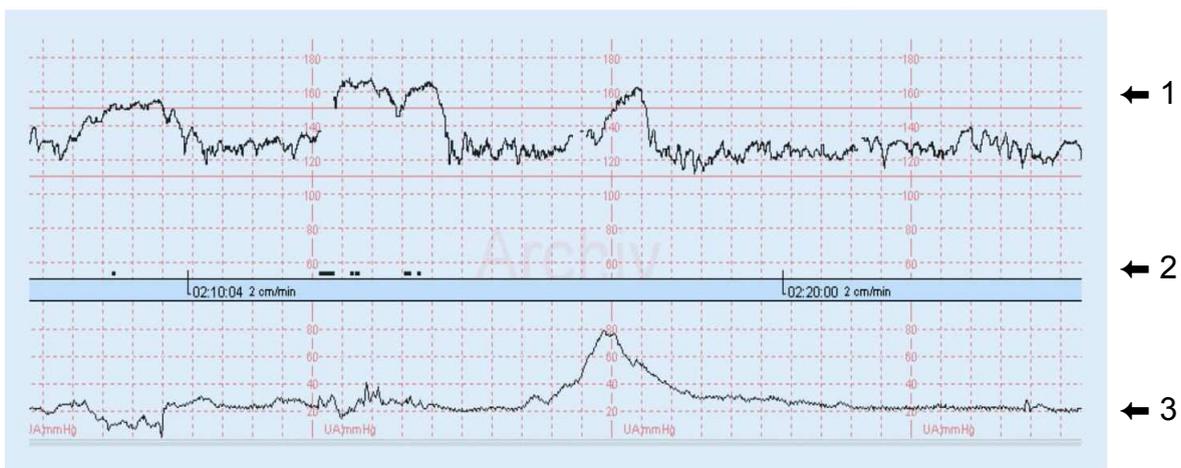
### 2.1 Kardiotokographie - Prinzip der Methode

Die Kardiotokographie ist eine Technik zur simultanen Abbildung der fetalen Herzaktion und der uterinen Kontraktionen. Eine wichtige zusätzliche Komponente ist die mütterliche Herzaktion. Moderne CTG-Geräte verfügen neben der Möglichkeit zur Erfassung der fetalen Herztöne über eine Funktion zur Überwachung der fetalen Bewegungen. In diesem Fall wird der Begriff "Kineto-Kardiotokographie" verwendet. Es werden in der Regel zwei Sensoren auf dem Abdomen der Schwangeren platziert. Der erste Sensor wird in der Nähe des fetalen Herzens positioniert. Er enthält spezielle Piezokristalle, die hochfrequente Schallwellen erzeugen, wenn eine Wechselspannung an sie angelegt wird. Wenn eine Ultraschallwelle auf dem Weg ins Gewebe eine bewegte Oberfläche trifft, reflektiert sie und verändert ihre Frequenz. Das Phänomen wird als Doppler-Effekt bezeichnet. Die Frequenzverschiebung wird durch die Piezokristallen detektiert; sie dienen nicht nur als Sender, sondern auch als Empfänger der Schallwellen. Anhand von Angaben über die Frequenzverschiebung der Schallwellen kann die fetale Herzfrequenz über die Formel "fetale Herzfrequenz = 60 Sekunden / Periodendauer (s)" berechnet werden. Durch die gleiche Sonde ist es möglich, die fetalen Bewegungen aufzuzeichnen. (Götz et al. 2015). Der zweite Sensor ist ein Drucktransducer, der die Frequenz der uterinen Kontraktionen registriert. Durch eine Modifikation des Transducers kann die Funktion um eine mütterliche Pulsaufzeichnung erweitert werden (Dongus 2011). Im Fall einer nicht ausreichenden Qualität des Signals besteht die Möglichkeit einer internen Ableitung der fetalen Herzaktion durch eine Kopfschwartenelektrode. Es entsteht dadurch ein geringes zusätzliches Infektionsrisiko und aus diesem Grund wird diese Elektrode nur nach Indikation eingesetzt.



**Abbildung 2: Schematische Darstellung der Kardiotokographie mit Positionierung der Sonden am Abdomen der Patientin (aus Barton Health)**

Alle relevanten Parameter werden auf einem Druckpapier erfasst und mindestens 30 Minuten abgeleitet.



**Abbildung 3: Beispiel eines Kardiotokogramms: 1. fetale Herzaktion, 2. fetale Bewegungen, 3.uterine Kontraktionen (Hösli et al. 2017)**

Ein repräsentatives Beispiel der CTG-Messung ist in der Abbildung 3 dargestellt. Die Papiervorschubgeschwindigkeit beträgt in den meisten geburtshilflichen Einrichtungen 1cm/min, sodass eine Standardaufzeichnung ca. 30 Zentimeter Papier verbraucht.

## 2.2 Auswertung des CTG nach den FIGO-Kriterien

In Deutschland wird die standardisierte Auswertung des CTG anhand der Kriterien der International Federation of Gynecology and Obstetrics (FIGO) empfohlen (Ayres-de-Campos et al.2015). Die Grundlagen der Kriterien sind:

1. Baseline-basale fetale Herzfrequenz. Die Einheit lautet Schläge pro Minute (SpM) und der Normbereich beträgt 110-160 SpM
2. Oszillation (Variabilität)-die Bandbreite der Schwankungen in der fetalen Herzaktion. Als normal gilt ein Bereich zwischen 5 und 25 SpM
3. Dezelerationen-Abfall der fetalen Herzfrequenz um mindestens 15 SpM innerhalb von 15 Sekunden. Die Dezelerationen werden nach Ursache und Bezug zu den uterinen Kontraktionen in drei Subtypen aufgeteilt
4. frühe Dezelerationen-erscheinen simultan zu den uterinen Kontraktionen und sind meistens Zeichen einer Kompression des fetalen Kopfes
5. variable Dezelerationen-treten unabhängig von uterinen Kontraktionen auf und werden als Reaktion auf Nabelschnurkompression bezeichnet
6. späte Dezelerationen-beginnen innerhalb von 20 Sekunden nach der Wehe und deuten auf fetale Hypoxie hin
7. Akzelerationen-Anstieg der fetalen Herzfrequenz um 15 SpM für mindestens 15 Sekunden und maximal 10 Minuten. Die Akzelerationen haben in den letzten Jahren an Bedeutung verloren. Die Abwesenheit von Akzelerationen bei sonst normalem Kardiotokogramm kann ein Zeichen des tiefen fetalen Schlafes sein und wird nicht mehr als "pathologisch" bewertet

	<b>normal</b>	<b>suspekt</b>	<b>pathologisch</b>
Baseline	110–160 SpM	Mindestens eines der normalen Kriterien fehlt, ohne dass weitere pathologische Zusatzkriterien vorliegen	<100 SpM
Variabilität	5–25 SpM		Reduzierte Variabilität während > 50 Min., erhöhte Variabilität während > 30 Min. oder sinusoidales Muster während > 30 Min.
Dezelerationen	Keine repetitiven Dezelerationen*		Wiederholte späte oder prolongierte Dezelerationen während > 30 Min. oder > 20 min bei reduzierter Variabilität oder eine prolongierte Dezeleration > 5 Min.
Interpretation	Keine Hypoxie/Azidose	Niedriges Risiko für Hypoxie/Azidose	Hohes Risiko für Hypoxie/Azidose

**Tabelle 1: Tabellarische Darstellung der FIGO-Kriterien. \*Repetitive Dezelerationen treten bei mehr als 50 % aller Uteruskontraktionen auf**

Bei einem Signalverlust von über 15 % (Schneider et al. 2016) gilt das Kardiotokogramm als nicht auswertbar. Die Auswertung nach FIGO-Kriterien wird u.a. in der aktuellen S3 Leitlinie “Vaginale Geburt am Termin” empfohlen (Abou-Dakn et al. 2020) und stellt aktuell den Goldstandard in der geburtshilflichen Praxis dar.

### **2.3 Studiendesign und ethische Aspekte**

Der Konzeption der Studie ging eine Literaturrecherche voraus. Nach Sichtung der Literatur wurden die Fallzahl sowie die Ein- und Ausschlusskriterien für die Studie festgelegt. Die Datenbankstruktur wurde in CTG-Daten im CTS-Format und Patientendaten in tabellarischer Form aufgeteilt. Die Patientendaten aus der klinischen Dokumentation sollten nach allgemeinen Informationen mit maternalen Risikoeigenschaften, fetalen Merkmalen und intrapartalen Informationen klassifiziert werden.

Nach dem kompletten Aufbau der Datenbank (REDCap-Datenbank) wurde die Phase der manuellen Qualitätssicherung geplant. Die auf diese Weise vorbereiteten Daten sollten statistisch und mit algorithmischen Methoden untersucht werden.

Vor Beginn der Studie wurde der Prüfplan zur Beratung der Ethikkommission an der medizinischen Fakultät der Eberhard-Karls-Universität in Tübingen vorgelegt. Im Juli 2020 stellte die Kommission ein positives Ethikvotum aus (Nr. 541/2020BO).

Die Analyse der Daten erfolgte retrospektiv anhand der klinisch-geburtshilflichen Dokumentation des Tübinger Perinatalzentrums. Die Untersuchungen nahmen keinerlei Einfluss auf das klinische Management. Eine retrospektive Aufklärung der Gebärenden über die Studie war nicht möglich und es wurde im Einvernehmen mit der Ethikkommission darauf verzichtet.

Ein Verarbeitungsverzeichnis wurde entsprechend der europäischen Datenschutz-Grundverordnung (DSGVO) angelegt.

## **2.4 Patientenkollektiv**

### **2.4.1 Ein- und Ausschlusskriterien**

Die Studienpopulation dieser Arbeit enthielt vitale Einlingsschwangerschaften ab der 24+0 SSW. Die Patientendokumentation aus der Universitäts-Frauenklinik Tübingen stammte aus den Jahren 2016-2021. Folgende Ausschlusskriterien wurden festgelegt:

- unvollständige geburtshilfliche Anamnese
- Signalverlustrate über 20 % in vorgeburtlichem CTG (30 Minuten vor der Geburt)
- unzureichende/fragliche Dokumentation des arteriellen Nabelschnur-pH-Werts

### **2.4.2 Fallzahl und Qualitätssicherung**

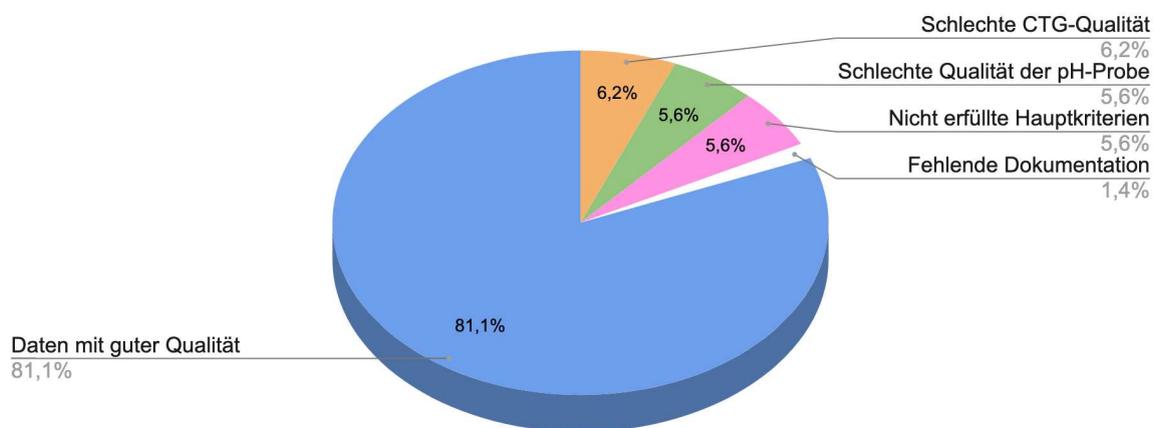
Die Fallzahl wurde als Kompromiss zwischen KI-Methoden und der geplanten manuellen Qualitätssicherung der Daten anfänglich mit mindestens 1100 geplant.

In der ersten Phase der Studie wurde die Dokumentation aus den Jahren 2018-2020 (1097 Fälle) gesichtet. Im Laufe der finalen Experimente, bei neu entstandenen Fragestellungen, wurden zusätzlich 34 Fälle aus den Jahren 2015-2017 und 2021 mit kritischen arteriellen pH-Werten eingeschlossen. Die endgültige Fallzahl der Studie betrug damit 1131.

Es erfolgte eine aufwendige manuelle Qualitätssicherung der automatisch

exportierten Daten. Die manuelle Methode ermöglichte eine detaillierte Einsicht in das Studienmaterial und eine Identifizierung der kritischen Probleme der automatisierten Datenextraktion. Es konnte dadurch eine effiziente automatische Qualitätssicherung beim zukünftigen Aufbau der Datenbank projiziert werden.

Es wurden anfänglich 1394 Fälle, sowohl Patientendaten als auch CTG-Abschnitte, automatisch exportiert. 81.1% der Primärdaten wurden in die Analyse eingeschlossen, 18.9 % erfüllten die Qualitätskriterien nicht.



**Abbildung 4: Ergebnisse der Qualitätssicherung: blaues Feld - gute Datenqualität. Oranges, grünes, violette und weißes Feld - Daten je nach Ausschlussgrund**

In der Gruppe der ausgeschlossenen Fälle war der häufigste Grund mit 6,2 % die schlechte Qualität des letzten 30-minütigen CTG-Ausschnitts. Bei einem Signalverlust von über 20% oder einer Dauer der Aufzeichnung unter 30 Minuten wurde das Kardiotokogramm als "schlecht auswertbar" klassifiziert.

In der Literatur wird als Qualitätsindikator ein Signalverlust von unter 15 % angegeben (Lasch et al. 2017). Die Erhöhung der akzeptablen Signalverlustrate erfolgte aufgrund der erschwerten Ableitung der fetalen Herztöne in den letzten 30 Minuten der Geburt (z.B. starke Geburtsarbeit, zahlreiche Bewegungsartefakte, Kind tief im Geburtsweg eingestellt). Die kontinuierliche Überwachung des Fetus in der letzten Phase der Entbindung wird durch die hohe Schmerzintensität dieser Periode beeinträchtigt.

Der zweite und der dritte Ausschlussgrund kamen mit 5.6% in gleicher Häufigkeit vor.

Bei einer schlechten Qualität der Nabelschnur-pH-Probe handelte es sich um eine Abnahme venösen anstatt arteriellen Blutes oder um eine Mischung davon. Bei den nicht erfüllten Hauptkriterien handelte es sich um Gemini-Graviditäten und avitale Schwangerschaften. Die fehlende Dokumentation, die bei 1.4 % der Daten vorkam, lag in den meisten Fällen in einer Sperrung des Zugangs zur Patientenakte für Unbefugte begründet.

### 2.4.3 Maternale Charakteristika und Risikoprofil

Aus den maternalen Daten wurden die allgemeinen Patientendaten und die Informationen aus dem geburtshilflichen Risikoprofil in die Analyse eingeschlossen. Die allgemeinen Patientendaten beinhalteten Alter, Gravidität, Parität und BMI zum Zeitpunkt der Entbindung.

Das anamnestische Risikoprofil ist in Tabelle 2 dargestellt.

Risikoeigenschaft	Kommentar
Diabetes mellitus	Vorkommen: ja oder nein (präexistente und Gestationsdiabetes)
Arterielle Hypertonie	1) schwangerschaftsinduzierte Hypertonie 2) Präeklampsie 3) HELLP-Syndrom 4) präexistente Hypertonie
Z.n. uteriner OP	1) Z.n. Sectio 2) Z.n. Kürettage oder Hysteroskopie 3) Z.n. Konisation 4) Z.n. Myomenukleation
Z.n. SA oder IUFT	Spätabort - 13+0 bis 23+6 SSW IUFT - ab 24+0 SSW
Z.n. Infertilitätsbehandlung	ICSI oder InVitro
Nikotinabusus	Anamnese: ja oder nein

**Tabelle 2: Mütterliches Risikoprofil - Kriterien und Erläuterungen**

Ausgewählt wurden die häufigsten Krankheiten bzw. anamnestischen Faktoren, die aus klinischer Sicht die größte Rolle in der Betreuung der Schwangeren spielen.

#### 2.4.4 Fetale Merkmale

Eine der wichtigsten fetalen Eigenschaften, die zur Risikoeinschätzung herangezogen werden, sind Gestationsalter, fetale Biometrie und fetaler/fetomaternaler Farbdoppler.

Die sonographischen Routineuntersuchungen der Schwangeren wurden bei stationärer Aufnahme im Tübinger Perinatalzentrum durchgeführt. Der Ultraschall im Studienkollektiv wurde mit ViewPoint Software Version 5 der Firma GE Healthcare dokumentiert. Die Messungen (fetale Biometrie, Fruchtwassermenge, Doppler, etc.) stammen aus einem Zeitraum von maximal 2 Wochen vor Entbindung.

Ein Überblick der gemessenen Werte zeigt Tabelle 3.

Messung	Kommentar	NB der ausgewählten Parameter
Fetales Gewicht	In Gramm und Perzentilen	Pc: 10 bis 90
Fetale Parameter	1) Abdomenumfang (AU) 1) biparietaler Durchmesser (BPD) 2) Femurlänge (FL)	Pc: 10 bis 90
Fetales Doppler	1) A.umbilicalis PI (AU) 2) A.cerebri media PI (ACM)	< 95 Pc > 5 (10) Pc
Fruchtwassermenge	- Amniotic Fluid Index (AFI) - Single Depot (SD)	6-16 cm 2-6 cm

**Tabelle 3: Fetale Charakteristika. Pc-Perzentil, PI-Pulsatilitätsindex, einer der Hauptparameter des Gefäßwiderstandes**

Die fetale Biometrie war Bestandteil jeder Aufnahmeuntersuchung. Fetaler Doppler und Fruchtwassermenge-Kontrolle erfolgten bei Terminüberschreitung oder bei V.a. Pathologie. Die Dokumentation der fetalen Parameter und des fetalen Dopplers verlief entsprechend in Millimeter und in Zahlen. Fetale Biometrie und in einem gewissen Maße fetaler Doppler sind abhängig vom Gestationsalter.

## 2.4.5 Intrapartale Daten

Die intrapartalen Daten zeigt Tabelle 4.

Parameter	Kommentar
Geburtsdatum und Zeitpunkt	
Art der Entbindung	1) spontane Geburt 2) primäre Sectio 3) sekundäre Sectio
Indikation zur sekundären Sectio	1) Geburtsstillstand 2) akute medizinische Indikation a) V.a. Uterusruptur b) vaginale Blutung c) drohende Asphyxie d) vorzeitige Wehentätigkeit 3) Risikoabwägung (BEL, Z.n. traumatisch erlebter Geburt, Z.n. Sectio) 4) andere medizinische Indikation (AIS, Präeklampsie, Z.n. Myomenukleation, Fehlbildung des Fetus)
Fruchtwasserfarbe	1) klar 2) grün 3) dick grün
Blutabnahme aus der Nabelschnur	- arterieller und venöser pH-Wert - Basenüberschuss
APGAR-Werte	aus Minute 1', 5' und 10' nach der Geburt

**Tabelle 4: Informationen über den peripartalen Zeitraum**

Ein arterieller und ein venöser pH-Wert sowie ein Basenüberschuss wurden kurz nach der Geburt aus der Nabelschnur-Arterie und -Vene abgenommen und als Kriterium der fetalen Hypoxie interpretiert. Nach Schneider et al. wird ein Nabelschnur-pH-Wert von unter 7.2 als Azidose bezeichnet, von einer schweren Azidose spricht man bei einem Wert von unter 7.0 (Schneider et al. 2016: 720).

pH-Wert	Azidität
≥7,30	Normaler Zustand
7,29-7,25	Reduzierter Zustand
7,24-7,20	Präazidose
7,19-7,15	Leichte Azidose
7,14—7,10	Mittelgradige Azidose
7,09-7,0	Fortgeschrittene Azidose
<7,0	Schwere Azidose

**Abbildung 5: Definition der fetalen Azidämie im Rahmen der Fetalblutanalyse (FBA) nach Saling**

Eine Geburtsazidose zeigt sich in der Regel als respiratorische oder als gemischte, respiratorisch-metabolische Störung (Perez et al. 2021). Die respiratorische Komponente ist typisch für ein akutes hypoxisches Ereignis der letzten Phase der Geburt. Ein länger anhaltender Sauerstoffmangel manifestiert sich durch den Übergang in eine gemischte, respiratorisch-metabolische Azidose (Bobrow et al. 1999). Der Basenüberschuss ist ein Parameter, der bezüglich weiterer Aussagen zu der metabolischen Komponente einer Azidose genutzt werden kann. Werte über 12 mmol/l korrelieren mit einer Verschlechterung des fetalen Outcome (Ross et al. 2002). Ein Wert gemäß der weltweit anerkannten APGAR-Skala wird von einem Arzt oder einer Hebamme in der 5. und der 10. Minute nach der Geburt vergeben. Er entspricht dem klinischen Zustand des Kindes.

Parameter	0 Punkte	1 Punkt	2 Punkte
<i>Atmung</i>	keine	unregelmäßig, flach	regelmäßig, Kind schreit
<i>Pulsfrequenz</i>	keine	unter 100/min	über 100/min
<i>Muskeltonus</i>	schlaff	partielle Beugung der Extremitäten	aktive Bewegungen
<i>Aussehen/Hautfarbe</i>	weiß/blau	Stamm rosig, Extremitäten blau	vollig rosig
<i>Reflexerregbarkeit</i>	keine	Grimassieren	kräftiges Schreien

**Abbildung 6: APGAR-Schema zur klinischen Beurteilung des Neugeborenen (Wert 8-10: gesundes Neugeborene, Wert 4-8: mittelschwere Adaptationsstörung, Wert < 4: schwere Asphyxie)**

## 2.5 Statistische Auswertung

### 2.5.1 Deskriptive Statistik und statistische Inferenz

Für die statistische Auswertung der Daten wurde das Programm IBM SPSS Statistics, Version 29.0 verwendet. Die Untersuchung der zwei unabhängigen Variablen erfolgte durch eine bivariate Analyse mit Hilfe des t-Tests. Ein Vergleich von mehr als zwei unabhängigen Variablen wurde mittels Varianzanalyse (ANOVA) durchgeführt.

In beiden Konstellationen wurde auf Normalverteilung (bei kleinen Proben) und auf Varianzhomogenität geprüft. Bei nicht erfüllten Voraussetzungen erfolgte die Analyse mit entsprechenden alternativen Testverfahren: dem Mann-Whitney-U-Test und der Welsch-ANOVA. Die deskriptive Statistik wurde in Form von Diagrammen und in tabellarischer Form mittels Google Docs Editor ausgearbeitet.

## **2.6 Methoden der künstlichen Intelligenz**

Der Aufbau der KI-Modelle basierte auf den Methoden der logistischen Regression und der faltungs-neuronalen Netzwerke. Die Studie beinhaltete zwei Datenarten: Patientinnendaten und Kardiotokogramme. Das gewünschte Outcome war die Einteilung der Daten in zwei Gruppen: mit fetalen arteriellen NS-pH-Wert  $\leq 7.2$  und  $> 7.2$ . Für diese Fragestellung wurden Methoden der Datenklassifikation angewandt.

Im ersten Schritt wurden die Patientendaten mittels logistischer Regression analysiert. Es ist eine der grundlegenden Methoden der künstlichen Intelligenz, die für die binäre Klassifizierungsprobleme verwendet wurde.

Im zweiten Schritt wurden die Daten aus der Kardiotokographie mit faltungs-neuronalen Netzwerken erforscht. Die faltungs-neuronalen Netzwerke werden typischerweise in einem Trainingsprozess programmiert. Die Daten wurden in dieser Arbeit in zwei Gruppen aufgeteilt: Trainingsdaten und Testdaten. An den Trainingsdaten wurde gelernt und an den Testdaten wurde das Gelernte geprüft. In einem Lernprozess wurden die Eingangsdaten ohne Outcome durch das Netz klassifiziert nach einem der gewünschten Bereiche (hier:  $\leq$  oder  $> 7.2$ ). Der Lernalgorithmus verbesserte andauernd diese Klassifikationsversuche, bis die maximale Leistungsfähigkeit des Algorithmus erreicht wurde. Für das Training eines Algorithmus wurden Tausende Versuche benötigt. Das trainierte Modell wurde zur Qualitätskontrolle auf die ihm bekannten Testdaten angesetzt. Im letzten Schritt der Analyse wurden die beiden Modelle zusammengefügt.

## **2.7 Aufbewahrung des Datenmaterials**

Die Datenerhebung erfolgte pseudonymisiert in Form einer Tabelle unter Angabe der Fallnummern der Patientinnen. Nach Ablauf der Studienlaufzeit wurden die Daten durch das Löschen der hinterlegten Fallnummern vollständig anonymisiert.

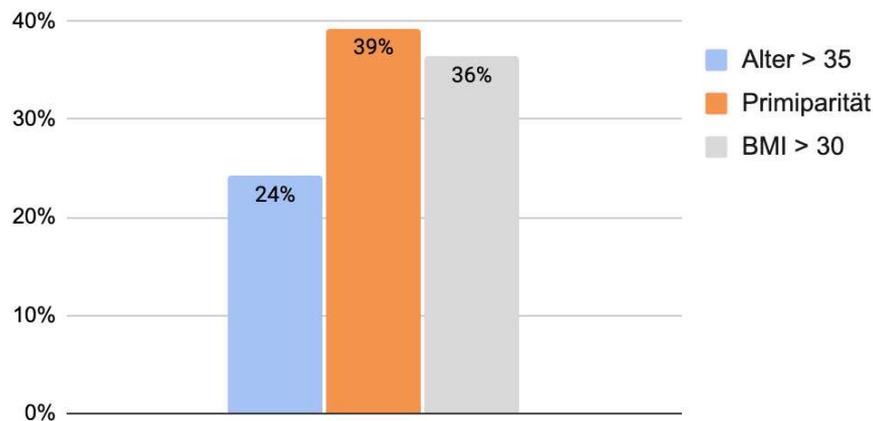
Die anonymisierten Daten wurden in eine REDCap-Datenbank übertragen. Den Zugriff auf die o.a. Daten erhielten ausschließlich mit der Datenerhebung und -auswertung betraute Wissenschaftler/-innen.

## 3 Ergebnisse

### 3.1 Charakteristika der Studienpopulation

#### 3.1.1 Subpopulation der Schwangeren

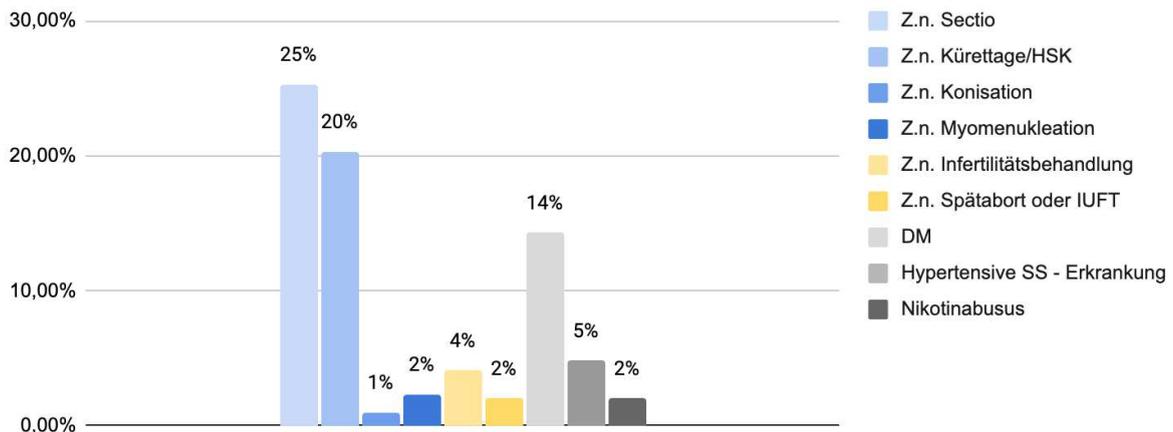
Das Durchschnittsalter des Studienkollektivs lag bei 32 Jahren. Im Schnitt hatte die Studienpatientin ein Kind geboren und ihr durchschnittlicher BMI betrug 29.



**Abbildung 7: Darstellung der schwangeren Population nach Alter, Parität und BMI, zum Zeitpunkt der Entbindung**

Abbildung 7 zeigt einen im Verhältnis hohen Anteil der Schwangeren, die über 35 Jahre alt sind, und per Definition zur Risikogruppe gehören (24 %), sowie einen hohen Anteil Primiparae (39 %). Das Tübinger Perinatalzentrum ist eine Einrichtung der höchsten Versorgungsstufe-Level 1, was das hohe Alter und den Anteil der Risikoschwangeren erklärt. In Betracht der BMI-Normen für die allgemeine Bevölkerung spricht man ab einem BMI über 30 von einer Adipositas.

Für die schwangere Population werden in der Regel der präpartale BMI und die Gewichtszunahme während der Schwangerschaft als Kriterien zur Kontrolle des Körpergewichts herangezogen. In Abbildung 7 wurde zur besseren Darstellung der peripartalen Situation der BMI vor der Geburt ausgewählt. Die BMI-Werte aus der Abbildung können nicht direkt mit Werten der allgemeinen Bevölkerungsgruppe verglichen werden.



**Abbildung 8: Maternales Risikoprofil**

In Abbildung 8 ist das maternale Risikoprofil nach Anzahl der Diagnosen dargestellt. Der häufigste Befund war Z.n. Sectio mit 25%; in 4 % der Fälle waren es zwei oder mehr vorangegangene Kaiserschnitte. Die zweithäufigste Diagnose (20%) war der Z.n. Kürettage. Auch hier hatten wiederum vier Prozent der Frauen zwei oder mehr Kürettagen erhalten. Erwähnenswert ist der mit 14% relativ hohe Anteil Diabetikerinnen, worunter sowohl Gestationsdiabetes als auch präkonzeptioneller Diabetes mellitus zusammengefasst werden. Von den hypertensiven Schwangerschaftserkrankungen waren mit 1.7 % die schwangerschaftsinduzierte Hypertonie, mit 1.3 % Präeklampsie, mit 0.7 % das HELLP-Syndrom und mit 1.1 % eine präexistente Hypertonie vertreten. Nikotinabusus wurde insgesamt in 2.3 % der Fälle vermerkt, was deutlich niedriger ist, als die in der Fachliteratur angegebenen Fallzahlen (Lange et al. 2018).

### 3.1.2 Fetales Kollektiv

Wie Tabelle 5 zeigt, betrug das durchschnittliche fetale Alter zum Zeitpunkt der Geburt 274 Tage.

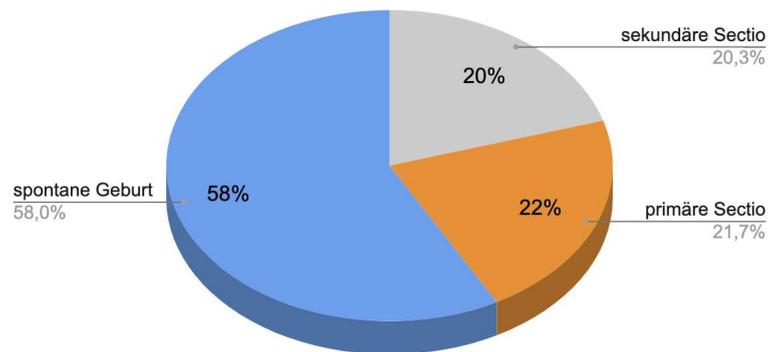
<b>Faktor</b>	<b>M</b>	<b>SA</b>	<b>%</b>
<b>Gestationsalter</b>	274	14.6	
GA > 41+2 SSW			4
GA < 37 SSW			9
<b>Fetales Gewicht (g)</b>	3259	616	
>4000			6
<2500			8
<b>Fetales Gewicht (Pc)</b>	44.2	29.1	
>90			7
<10			10
<b>A.umbilicalis PI</b>	0.9	0.3	
<b>A.cerebri media PI</b>	1.5	0.3	
<b>AFI</b>	9.9	4.9	
<b>SD</b>	4.5	2.3	

**Tabelle 5: Gestationsalter zum Zeitpunkt der Geburt, fetale Biometrie und Farbdoppler aus der Aufnahmeuntersuchung-maximal 2 Wochen vor der Entbindung. M-Mittelwert, SA-Standardabweichung, PI-Pulsatility Index, Pc-Perzentile, AFI-Amniotic Fluid Index, SD-Single Depot**

Die Zahl der Frühgeburten war etwa doppelt so hoch wie die der Übertragungen über 41+3 SSW 9% gegenüber 4%. Es fällt der hohe Anteil der hypotrophen Feten auf. In 10% der Fälle lag das fetale Schätzwert unter der 10. Perzentile. Eine Hypertrophie wurde in 7% der Fälle festgestellt. Insgesamt hatte fast jeder fünfte Fetus eine abnorme Gewichtsschätzung, was auf die hohe Versorgungsstufe (Konzentration von Risikoschwangerschaften) des Tübinger Perinatalzentrums zurückzuführen ist.

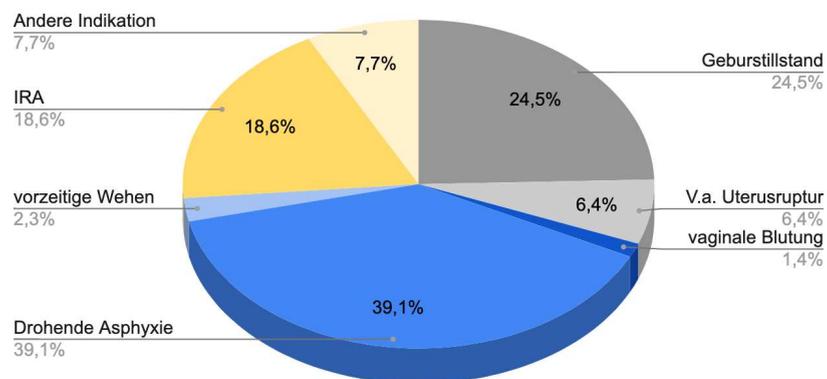
### **3.1.3 Das Ereignis der Geburt**

Bei fast 60% der Geburten handelte es sich um Spontangeburt. Die Sectio-Rate betrug insgesamt 42.3 %, was über dem bundesweiten Durchschnitt liegt. Die genaue Aufteilung ist in Abbildung 9 dargestellt.



**Abbildung 9: Aufteilung der Entbindungsmodi**

Der häufigste Grund für eine sekundäre Sectio caesarea war die drohende Asphyxie, der zweithäufigste ein Geburtsstillstand. In der Gruppe der Patientinnen, bei denen nach einer Risikoabwägung eine Indikation zur Sectio gestellt wurde, ging es vor allem um das Vorziehen eines geplanten Sectio-Termins bei Geburtsbeginn. Die genauen Werte sind aus Abbildung 10 ersichtlich.



**Abbildung 10: Aufteilung der sekundären Kaiserschnitte nach Indikation**

Am Anfang des Projektes bestand die Studienpopulation aus 1097 Fällen, im weiteren Verlauf wurden zusätzlich 34 Fälle mit besonders schlechten (< 7.0) arteriellen pH-Werten aus den Jahren 2015-2017 und 2021 eingeschlossen.

<b>Faktor</b>	<b>M</b>	<b>SD</b>	<b>% &lt; 7.05</b>	<b>% &lt; 7.2</b>	<b>% &lt; 7</b>
<b>pH(A)</b>	7.24	0.07	0.6	24	
pH(B)	6.93	0.06	100	100	
pH(A+B)	7.23	0.08	3.6	27	
<b>APGAR 1' (A)</b>	8.4	1.4			
APGAR 1' (B)	4.2	2.5			
APGAR 1' (A+B)	8.3	1.6			
<b>APGAR 5' (A)</b>	9.3	0.9			1
APGAR 5' (B)	6.7	2.0			41
APGAR 5' (A+B)	9.3	1.0			3
<b>APGAR 10' (A)</b>	9.7	0.6			
APGAR 10' (B)	7.8	2.0			
APGAR 10' (A+B)	9.7	0.7			

**Tabelle 6: Charakteristika der Outcome-Daten. A-primäre Studienpopulation (1097 Fälle), B-zusätzliche Fälle mit kindlichem arteriellem pH-Wert < 7.0 (34 Fälle)**

Die primäre Studienpopulation (A) beinhaltete eine sehr niedrige Quote der schwerer peripartaler Azidosen - 0.2 %. Mit Hinzunahme der zusätzlichen Fälle (B) stieg der Anteil auf 3 %.

24% der peripartalen arteriellen pH-Werte in der Gruppe der primären Studienpopulation (27% nach Einschluss der zusätzlichen 34 Fälle) waren niedriger als 7.2, was auch als Cut-off-Wert für den Aufbau des KI-Algorithmus eingesetzt wurde.

## 3.2 Patientendaten und deren Zusammenhang mit dem arteriellen Nabelschnur-pH-Wert

### 3.2.1 Gesamtpopulation

Bisherige Arbeiten zum Thema CTG-Auswertung mit Methoden der künstlichen Intelligenz waren allein auf die Kardiotokographie fokussiert. Die Integration der Patientendaten in eine solche Analyse könnte die Effektivität der Algorithmen verbessern. Um diese Aussage zu überprüfen, wurden die Patientendaten in Bezug auf den arteriellen Nabelschnur-pH-Wert analysiert. Tabelle 7 stellt die Analyse der Gesamtpopulation dar.

Faktor	Nummer	Durchschnitt	SA	p-value
<b>Maternales Alter</b>				
< 35	752	7.24	0.07	0.12
≥ 35	345	7.24	0.07	
<b>Parität</b>				
0	445	7.24	0.07	0.6
≥ 1	650	7.24	0.07	
<b>Gestationsalter</b>				
37+0 - 41+3	947	7.24	0.07	
vs <37	103	7.27	0.06	< 0.001
vs > 41+3	47	7.22	0.07	0.18
<b>AP Primipara (Minuten)</b>				
< 120	146	7.21	0.06	0.6
≥ 120	42	7.22	0.06	
<b>AP Multipara (Minuten)</b>				
< 60	298	7.21	0.06	0.5
≥ 60	153	7.22	0.07	
<b>Z.n. Sectio</b>				
0	867	7.23	0.07	< 0.001
1	185	7.25	0.06	

<b>Faktor</b>	<b>Nummer</b>	<b>Durchschnitt</b>	<b>SA</b>	<b>p-value</b>
≥ 2	45	7.29	0.03	
<b>Z.n. Abrasio</b>				
0	907	7.24	0.07	0.3
1	143	7.24	0.07	
≥ 2	47	7.25	0.07	
<b>Infertilitätsbehandlung</b>				
Ja	45	7.25	0.07	0.2
Nein	1052	7.24	0.07	
<b>DM</b>				
Ja	159	7.24	0.08	0.3
Nein	938	7.24	0.07	
<b>Hypertensive SS-Erkrankungen</b>				
Nein	1045	7.24	0.07	
vs SIH	18	7.26	0.04	0.3
vs Präeklampsie	15	7.29	0.05	0.001
vs HELLP-Syndrom	8	7.30	0.02	0.002
vs präexistente aHT	11	7.26	0.04	0.16
<b>Nikotinabusus</b>				
Ja	25	7.27	0.05	0.02
Nein	1072	7.24	0.07	
<b>BMI</b>				
< 30	677	7.23	0.07	< 0.001
≥ 30	405	7.25	0.06	
<b>Fetales Schätzwicht (g)</b>				
2500-4000	765	7.24	0.07	
vs < 2500	94	7.28	0.06	< 0.001
vs > 4000	63	7.24	0.07	0.965
<b>Fetales Schätzwicht (Pc)</b>				
10 - 90	730	7.24	0.07	

<b>Faktor</b>	<b>Nummer</b>	<b>Durchschnitt</b>	<b>SA</b>	<b>p-value</b>
vs < 10	115	7.25	0.07	0.05
vs > 90	75	7.25	0.06	0.2
<b>Doppler A. umb. PI</b>				
< 90	119	7.25	0.07	0.13
≥ 90	784	7.24	0.07	

**Tabelle 7: Analyse der Patientendaten in Bezug auf arterielle Nabelschnur-pH-Werte in der Gesamtpopulation (1097 Fälle). P-value unter 0.05 weist auf statistische Signifikanz hin**

Es fällt auf, dass die Faktoren, die mit einer statistisch signifikanten pH-Veränderung einhergehen, für die Subgruppen der Geburten bezeichnend sind, die überwiegend Kaiserschnitte beinhalteten. Die Sectiorate in diesen Subgruppen beträgt:

- 65 % bei Geburten vor der 37. SSW
- 74 % bei Geburten in der Gruppe der Schwangeren mit Z.n. 1x Sectio
- 77 % bei Geburten mit Diagnose einer hypertensiven SS-Erkrankung bei der Mutter
- 54 % bei Geburten in der Gruppe der Patientinnen mit BMI ≥ 30
- 72 % bei Geburten mit einem Schätzwert des Fetus von unter 2.5 kg

In der Analyse der Gesamtpopulation zeigte sich, dass bei den Kindern der Patientinnen aus der Risikogruppe (mit einem der folgenden Faktoren: Frühgeburt, Z.n. Sectio caesarea, hypertensive SS-Erkrankung, BMI > 30 oder Schätzwert unter 2.5 kg) durchschnittlich ein höherer arterieller Nabelschnur-pH-Wert gemessen wurde.

Die beiden Entbindungsmodi unterschieden sich in Bezug auf den arteriellen Nabelschnur-pH-Wert signifikant. Der durchschnittliche Wert betrug 7.27 in der Sectio-Gruppe und 7.21 in der Gruppe der spontanen Geburten, was ein Grund für die genannten nicht intuitiven Resultate sein könnte.

### 3.2.2 Subpopulation der spontanen Entbindungen

Es wurde die gleiche Analyse in der Subpopulation der spontanen Geburten durchgeführt. Diese Subpopulation scheint den Geburtsstress und die darauf folgende Senkung des arteriellen Nabelschnur-pH-Werts besser abzubilden.

Faktor	Nummer	Durchschnitt	SA	p-value	*p-value
<b>Maternales Alter</b>					
< 35	464	7.21	0.07	0.6	0.12
≥ 35	172	7.21	0.06		
<b>Parität</b>					
0	230	7.21	0.06	0.13	0.6
≥ 1	405	7.22	0.07		
<b>Gestationsalter</b>					
37+0 - 41+3	572	7.21	0.06		
vs <37	36	7.25	0.07	< 0.001	<0.001
vs > 41+3	28	7.19	0.07	0.08	0.18
<b>Z.n. Sectio</b>					
0	586	7.22	0.06	0.04	<0.001
1	48	7.20	0.07		
≥ 2	(2)				
<b>Z.n. Abrasio</b>					
0	538	7.21	0.06	0.7	0.3
1	78	7.21	0.07		
≥ 2	26	7.21	0.07		
<b>Infertilitätsbehandlung</b>					
Ja	21	7.21	0.07	0.98	0.2
Nein	615	7.21	0.06		
<b>DM</b>					
Ja	72	7.22	0.07	0.79	0.3
Nein	564	7.21	0.06		

Faktor	Nummer	Durchschnitt	SA	p-value	*p-value
<b>Hypertensive SS-Erkrankungen</b>					
Nein	624	7.21	0.06		
vs SIH	8	7.25	0.05	0.16	0.3
vs Präeklampsie	3				0.001
vs HELLP-Syndrom	0				0.002
vs präexistente aHT	1				
<b>Nikotinabusus</b>					
Ja	9	7.24	0.06	0.2	0.02
Nein	627	7.21	0.06		
<b>BMI</b>					
< 30	441	7.21	0.06	0.4	< 0.001
≥ 30	185	7.22	0.07		
<b>Fetales Schätzwert (g)</b>					
2500-4000	440	7.21	0.06		
vs < 2500	26	7.25	0.08	0.04	<0.001
vs > 4000	29	7.2	0.06	0.24	0.965
<b>Fetales Schätzwert (Pc)</b>					
10 - 90	421	7.21	0.07		
vs < 10	51	7.21	0.06	0.8	0.05
vs > 90	21	7.21	0.05	0.8	0.2
<b>Doppler A. umb. PI (Pc)</b>					
< 90	44	7.21	0.06	0.4	0.13
≥ 90	444	7.22	0.06		

**Tabelle 8: Analyse der Patientendaten in Bezug auf arterielle Nabelschnur-pH-Werte in der Subpopulation der spontanen Geburten. P-value unter 0.05 weist auf statistische Signifikanz hin. \*p-value entspricht dem Wert aus der Gesamtpopulation (siehe Tabelle 7)**

Der Faktor "Z.n. Sectio" zeigte eine signifikante Rolle in der Senkung des arteriellen Nabelschnur-pH-Wertes (p value 0.04). Im Z.n. Re-Sectio wird in vielen Fällen von einer vaginalen Geburt abgeraten.

Ein Gestationsalter unter 37 SSW und ein Schätzwert unter 2500g waren Eigenschaften, die signifikanten Einfluss auf die Erhöhung des arteriellen Nabelschnur-pH-Wertes hatten. In den beiden Analysen der Gesamtpopulation und der Subgruppe der spontanen Geburten blieben der Zusammenhang und ihre Richtung signifikant (Analyse der Gesamtpopulation p value < 0.001, Analyse der Subgruppe von spontanen Geburten p-value 0.04).

BMI und Nikotinabusus hatten in der Subpopulation der spontanen Geburten im Unterschied zur Analyse der Gesamtpopulation keine Signifikanz im Bezug auf den Nabelschnur-pH-Wert mehr. Erwähnenswert ist die Tatsache, dass nur 9 Frauen (1.4%), die spontan entbunden haben, einen Nikotinabusus angegeben hatten. Dies ist ein deutlich niedrigeres Niveau als dies anhand der Daten der Fachliteratur zu erwarten gewesen wäre.

Eine Vergleichsanalyse bezüglich der hypertensiven SS-Erkrankungen konnte nicht durchgeführt werden. Die Mehrheit der Frauen erhielt eine geplante Sectio, die aufgrund potentieller Komplikationen für Mutter und Kind indiziert war. Die Anzahl der Patientinnen mit hypertensiven SS-Erkrankungen, die spontan entbunden haben, variierte in den verschiedenen Kategorien zwischen 1 und 8. Eine Übertragung des Termins über 41+3 SSW zeigte eine Tendenz (p-value 0.08) zur Senkung des arteriellen Nabelschnur-pH-Wertes.

### **3.3 Datenanalyse mit Algorithmen der künstlichen Intelligenz**

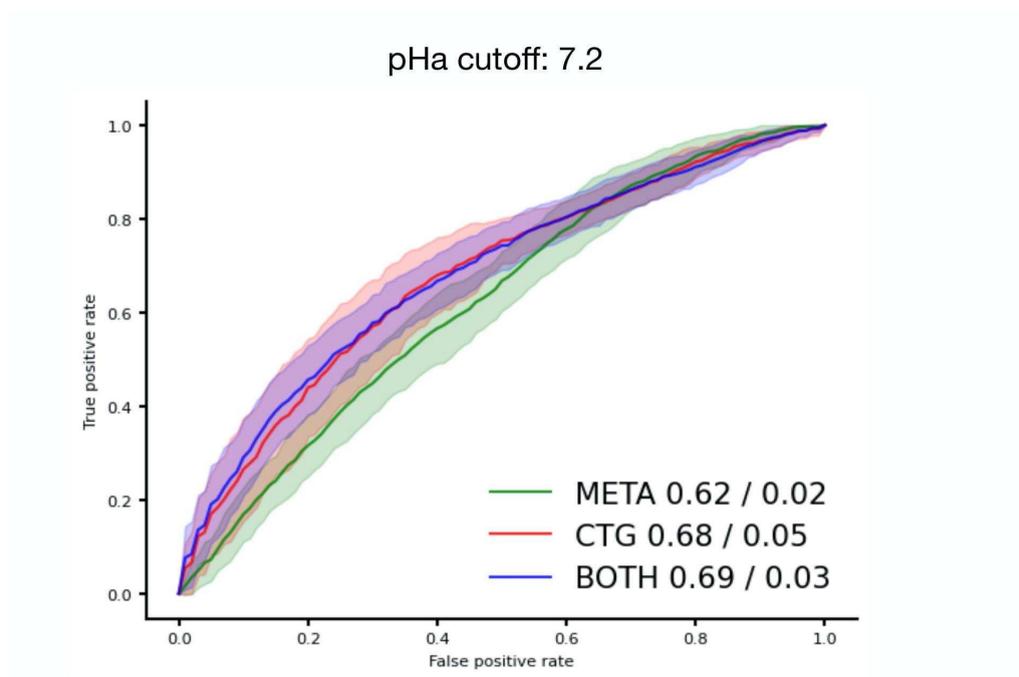
#### **3.3.1 Kardiotokogramme und Patientendata der Gesamtpopulation**

Für diese Phase der Studie wurde ein gefaltetes neuronales Netz konstruiert (engl. CNN - Convolutional Neural Network). Diese Art des neuronalen Netzes wird häufig im Bereich der Bildverarbeitung verwendet, eignet sich aber auch gut für die Analysen von Zeitreihendaten. Vorteile dieser Methode sind unter anderem eine gute Anwendungsmöglichkeit an sehr großen Datenbanken, hohe Effektivität in Projekten

ohne Feature-Engineering (ohne Vorverarbeitung der Rohdaten durch einen Experten) und Robustheit gegenüber Datenrauschen.

Die ersten Ergebnisse der Studie zeigt die Abbildung 11. Hier wurde die Leistungsfähigkeit des Algorithmus anhand einer so genannten ROC-Kurve (engl. Receiver Operating Curve) dargestellt. Die ROC-Kurve ist ein statistisches Verfahren, mit dem die Genauigkeit der Modellvorhersage in einer graphischen Form dargestellt wird.

Eine Kurve, die sich der linken oberen Ecke des Diagramms nähert, weist eine höhere Effektivität des Modells (hier: blaue Kurve) auf. Eine Kurve, die sich einem 45-Grad-Winkel nähert, zeigt eine schlechtere Leistungsfähigkeit (hier: grüne Kurve).



**Abbildung 11: META - Patientendaten, CTG - Daten aus der Kardiotokogramme, BOTH - Kombination der beiden Datenkategorien**

Die Vorhersage des arteriellen Nabelschnur-pH-Wertes betrug bei einem Cut-off-Wert des Nabelschnur-pH-Wertes von 7.2 alleine anhand der Daten aus den Kardiotokogrammen 0.68 AUC.

Die "Fläche unter der Kurve" (engl. AUC: Area Under the Curve) ist ein Parameter, der im Machine Learning als Maß der Effektivität der Datenklassifikation verwendet wird. Eine effektive Leistung beginnt bei AUC-Werten von über 0.7 (Nahm 2022). Die

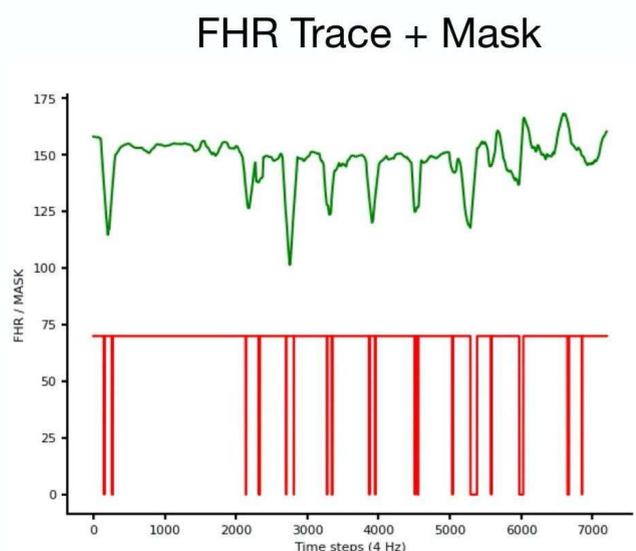
Eingabe der Patientendaten in das auf der Gesamtpopulation basierende Modell, erhöhte die Effektivität des Algorithmus lediglich von AUC 0.68 auf AOC 0.69.

### 3.3.2 Blackbox Effekt

Die Daten aus den Kardiotokogrammen wurden im CTS-Format gespeichert. Eine Standard-CTG-Aufnahme wird mit einer Frequenz von 4 Hz gemessen. Eine CTS-Datei enthält eine Liste der Messungen, die alle 250 Millisekunden erfolgen. Bei einem Signalverlust werden die Messungen als "0" dokumentiert.

Es besteht die Möglichkeit, die visuelle Qualitätssicherung der Daten durch einen automatischen Prozess zu ersetzen. Eine visuelle Qualitätssicherung entspricht aber dem klinischen Alltag. Bei einer automatischen Kontrolle würde die Zahl der Kardiotokogramme steigen, welche das Kriterium von weniger als 20 % Signalverlust nicht erfüllen.

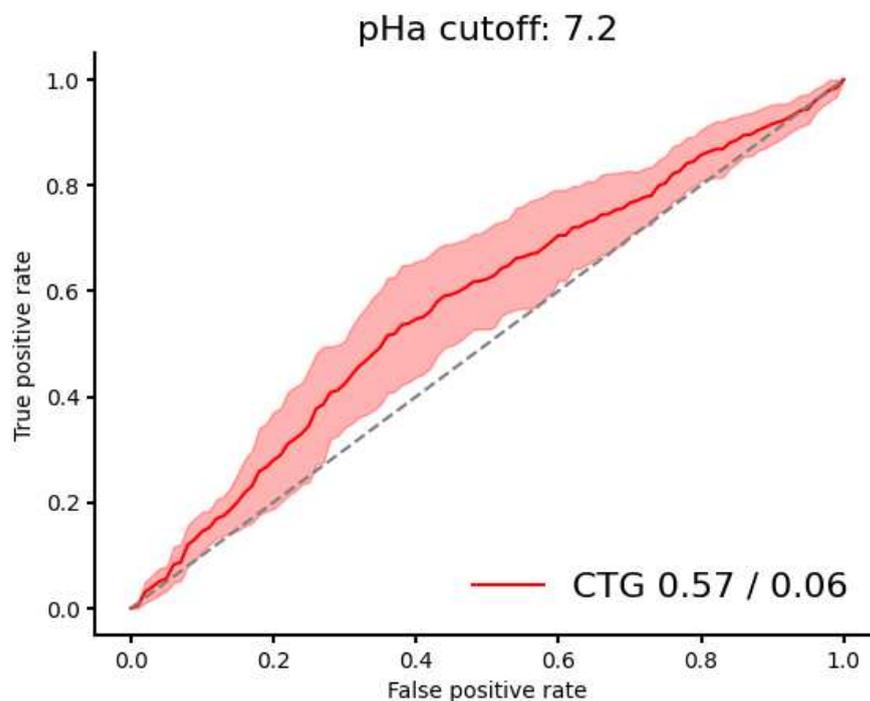
Bei der graphischen Darstellung der CTS-Dateien ist es in manchen Fällen ein sich wiederholendes Muster des Signalverlustes kurz vor der Entbindung aufgefallen. Es wurden dann alle Kardiotokogramme im Hinblick auf solche Signalverlust-Muster untersucht, und für jeden Fall wurde graphisch das Ausmaß der Signalverluste dargestellt. Diese Grafik wurde studienintern als "Maske" bezeichnet, ein Beispiel ist in Abbildung 12 zu sehen.



**Abbildung 12: Grüne Kurve: Herzaktion des Fetus, rote Kurve: Signalverlust**

Alle Kardiotokogramme wurden anhand ihrer Masken untersucht, und es konnte festgestellt werden, dass es in der Subpopulation der Kaiserschnitt-Geburten ein typisches Signalverlust-Muster gab. Dieses Muster entspricht der Pause in der CTG-Aufnahme während der präoperativen Vorbereitung der Patientin.

Im nächsten Schritt wurde das ursprüngliche KI-Modell an der Subpopulation der spontanen Geburten getestet. Die Ergebnisse sind in Abbildung 13 dargestellt und sollen im Folgenden erläutert werden.



**Abbildung 13: Effektivität des KI-Modells auf der Subgruppe der spontanen Geburten**

Die Effektivität des Algorithmus zur Vorhersage des Outcomes (Nabelschnur-pH-Wert kleiner oder größer als 7.2) beträgt in der Subpopulation der spontanen Geburten AUC 0.57. Für die künstlichen Intelligenz hat ein Lernprozess des Algorithmus nicht selten einen enorm komplizierten Ablauf. Sogar Datenwissenschaftler können nicht immer jedes Detail des Lernprozesses ihrer Algorithmen erklären. Diesen Effekt nennt man im Machine Learning den Blackbox-Effekt (Jiménez 2022).

Das Phänomen des Lernens anhand falscher Assoziationen wird in der Psychologie und der künstlichen Intelligenz als *Clever-Hans-Phänomen* bezeichnet (Hernández-Orallo 2019). Der "kluge Hans" war ein deutsches Pferd, das vor dem Ersten Weltkrieg lebte. Angeblich sollte es hochintelligent sein und die Fähigkeit zum Rechnen und Schreiben besitzen. Nach zahlreichen interdisziplinären Experimenten ergab sich, dass das Pferd die eigentlichen Fragen nicht verstand, sondern die Körpersprache des Menschen interpretierte und somit auf gestellte Aufgaben "richtig" reagieren konnte.

Das o.g. Phänomen wurde in unserer Studie beobachtet, indem das KI-Modell erlernt die "Maske" mit der Subpopulation Kaiserschnitt-Geburten und dadurch mit einem besseren Nabelschnur-pH-Wert zu assoziieren. Ein neuer, sich dynamisch entwickelnder Wissensbereich der KI-"Explainable AI"-beschäftigt sich damit, die Lernprozesse der Algorithmen zu verstehen und zu erklären. Bis dato stehen aber die meisten KI-Algorithmen mit ihrer imponierenden Rechenleistung nur in Kombination mit dem Blackbox-Effekt zur Verfügung.

### **3.3.3 Subpopulation der sekundären Kaiserschnitt-Geburten**

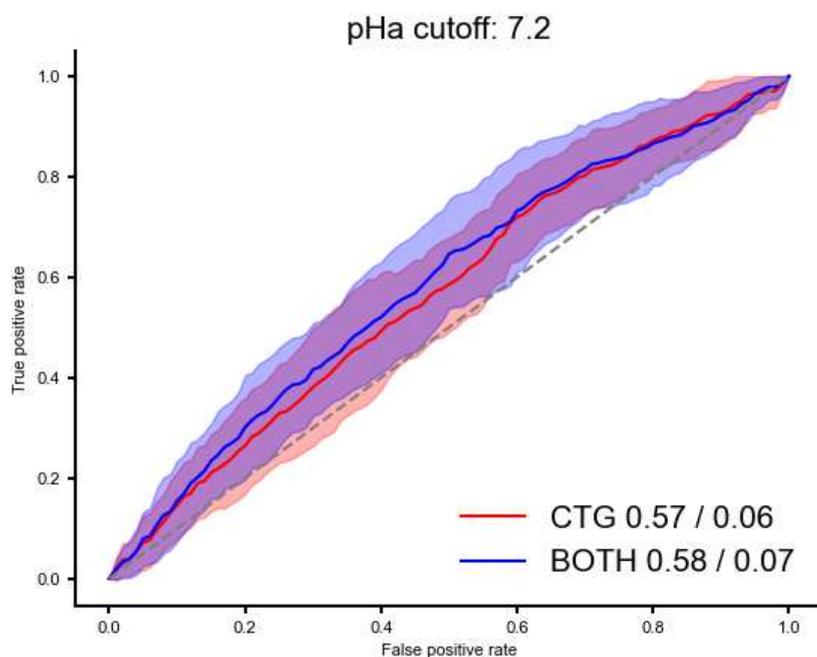
Im Laufe der Studie ergab sich die Fragestellung, ob die Gruppe von spontanen Geburten um die Fälle aus der Subgruppe der sekundären Kaiserschnitt-Geburten, die aufgrund der Diagnose "drohende Asphyxie" durchgeführt worden sind, erweitert werden könnte.

Die Zahl der spontanen Geburten im Tübinger Dataset betrug 636 Fälle, die Zahl der Kaiserschnitte mit der Indikationsstellung "drohende Asphyxie" betrug 86 Fälle. Der durchschnittliche pH-Wert in der Subpopulation der spontanen Geburten betrug 7.21, in der Subgruppe operativer Entbindungsmodi 7.25 und der Unterschied zwischen beiden Gruppen war signifikant ( $p\text{-value} < 0.001$ ). Dieses Ergebnis weist auf das Auftreten des Masken-Effektes hin. Die hier besprochene Subgruppe der sekundären Kaiserschnitt-Geburten konnte nicht an die endgültige Studienpopulation angeschlossen werden.

### 3.3.4 Kardiotokogramme und Patientendaten: Subpopulation der spontanen Geburten

Die statistische Analyse der Patientendaten in Bezug auf den arteriellen Nabelschnur-pH-Wert ergab unterschiedliche Ergebnisse für die Subpopulationen der Kaiserschnitt und der spontanen Geburten (SG). Der Einfluss der Risikofaktoren auf die Senkung des pH-Wertes in der Subgruppe der SG entsprach dem in der Literatur angegebenen und im klinischen Alltag bekannten biologischen Zusammenhang.

Es entstand die Fragestellung, ob die Eingabe der Patientendaten zu den Kardiotokogrammen die Leistungsfähigkeit des Algorithmus verbessert. Dementsprechend wurde ein weiteres Experiment durchgeführt. Die Ergebnisse sind in Abbildung 14 dargestellt.

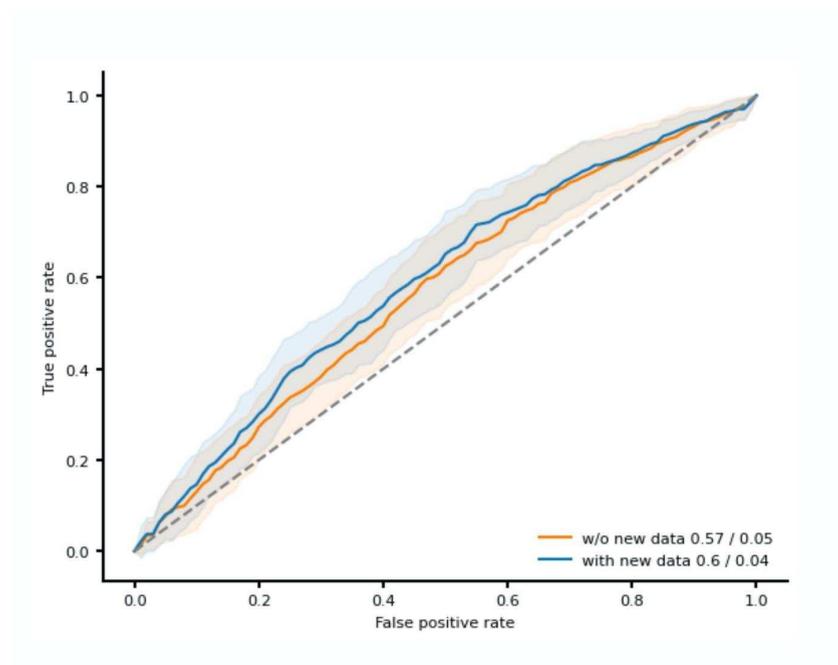


**Abbildung 14: CTG - Daten aus den Kardiotokogrammen, BOTH - Kombination der Kardiotokogramme mit Patientendaten**

Die Effektivität des Modells konnte nach Eingabe der Patientendaten von AUC 0.57 auf AUC 0.58 erhöht werden. Das Ergebnis war mit der Analyse der gesamten Studienpopulation vergleichbar (Erhöhung der Effektivität um 0.01 AUC).

### 3.3.5 Leistungsfähigkeit bei erhöhtem Anteil azidotischer Fällen

In der primären Studienpopulation kamen schwere Azidosen (pH-Wert < 7.05) mit einer Häufigkeit von 0.6 % vor, was mit dem in der Literatur angegebenen Wert von 0.68 % vergleichbar ist (Jonsson et al. 2009). Die Rate an schweren Azidosen (pH-Wert < 7.05) in der öffentlichen Datenbank der Universitätsklinik in Brno beträgt 7 % und in dem aus Oxford stammenden Dataset 4% (Arbeitsgruppe Petroziello et al.). Im Verlauf der Studie wurden zusätzlich 34 pathologische Fälle aus den Jahren 2015-2017 und 2021 eingeschlossen. Damit wurde der Anteil der schweren Azidosen auf 3 % erhöht.

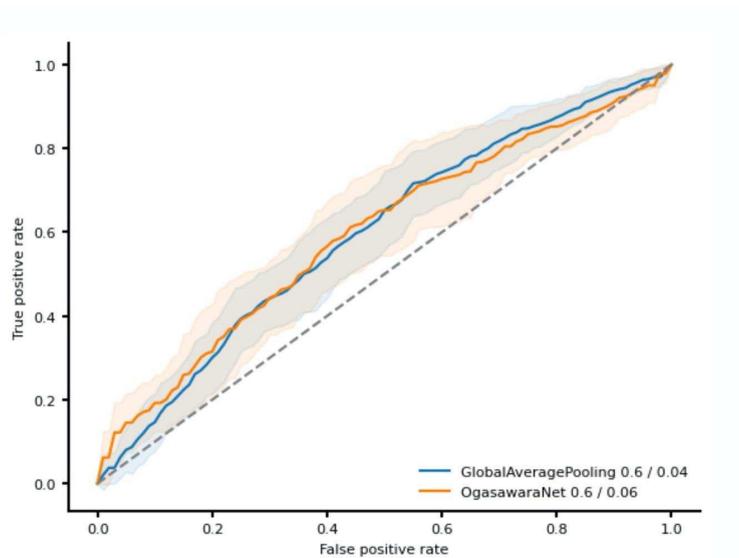


**Abbildung 15: Leistungsfähigkeit des Modells nach der Aufnahme von zusätzlichen pathologischen Fällen**

Die Effektivität des Algorithmus wurde von 0.57 auf 0.6 AUC erhöht.

Die Aufnahme der Fälle mit extremen Werten konnte demnach die Leistungsfähigkeit des Algorithmus erhöhen. Ob ein auf diesem Weg konstruiertes Modell die gleiche Effektivität an Echtzeitdaten zeigt, wird im Kapitel 4 dieser Arbeit diskutiert.

### 3.3.6 Re-Implementierung eines externen Modells auf eigenen Daten



**Abbildung 16: Effizienz des externen Modells auf eigenen Daten (1131 Fälle).  
GlobalAveragePooling: eigenes Modell, OgasawaraNet: externes Modell**

Als eines der letzten Experimente wurde die Re-Implementierung des Algorithmus aus der Arbeitsgruppe Ogasawara et al. auf eigenen Daten durchgeführt. Es wurde ein Ergebnis von AUC 0.6 erzielt, im Vergleich zu dem Ergebnis aus dem originalen Experiment mit AUC 0.73.

## 4 Diskussion

Die Überwachung der fetalen Herzaktion mittels Auskultation begleitet die Geburtshelfer seit über 120 Jahren und vermutlich sogar länger. Die Kardiotokographie (wie wir sie heute kennen) im Sinne der simultanen Herzton- und Wehentätigkeit-Überwachung mit gleichzeitiger graphischer Darstellung, ist seit über 50 Jahren bekannt und wird weltweit angewandt.

Im Hinblick auf die aktuelle diagnostische Stärke der Methode fallen sowohl die niedrige Spezifität, die hohe falsch positive Rate und die daraus folgende Steigerung der Sectorate als auch die diskussionsbedürftige Indikationsstellung für CTG-Anwendung im Niedrigrisiko-Kollektiv auf.

Die Studien, die eine antepartale kardiotokographische Überwachung mit keiner Überwachung vergleichen, existieren nicht und können aufgrund der rechtlich-ethischen Bedenken auch nicht geplant werden. Die wissenschaftliche Lage ist nicht zufriedenstellend und die erschwerte Durchführbarkeit neuer Studien mit diesen Fragestellungen im Schwangeren-Kollektiv (ethisches Dilemma) trägt dazu bei.

Es besteht ein Forschungsbedarf im Bereich der CTG-Auswertungssysteme. Eine mindestens ebenso gute Leistungsfähigkeit wie die ärztliche Bewertung (99% Sensitivität, 77% Spezifität), ist bei der niedrigen Quote der fachlichen Urteilübereinstimmung bezüglich der CTG-Interpretation (ca 48 %) anzustreben.

Von den auf den FIGO-Kriterien basierenden existierenden Auswertungssystemen fehlen entweder randomisierte prospektive Studien oder die Ergebnisse dieser Studie weisen auf keine signifikante Verbesserung des kindlichen Outcome hin.

Die Anwendung von Methoden der künstlichen Intelligenz hat das Potenzial einer Verbesserung der Ergebnisse bei ständig wachsenden Datenbanken. Die bestehenden Studien zeigen auf den ersten Blick vielversprechende Ergebnisse. Hier kann eine Interpretation der Resultate schwieriger als in der traditionellen Forschung sein. Doch wie in der traditionellen medizinischen Forschung ist nicht nur das Ergebnis, sondern auch die Studienplanung wichtig.

Ein wichtiger Aspekt von Studien zum Thema CTG-Beurteilung mit KI-Modellen ist der Aufbau des Algorithmus auf der internen Datenbank. Aufgrund der Datenschutzregelungen existieren in der Medizin nur wenige öffentliche Datenbanken. Die einzige frei zugängliche Datenbank, welche Kardiotokogramme

und Daten bezüglich des arteiellen Nabelschnur-pH-Wertes beinhaltet, ist eine Datenbank aus der Universitätsklinik in Brno (UTU-CHB). Der Aufbau einer eigenen Datenbank für das Training des KI-Modells ermöglicht die externe Validierung mindestens auf den klinischen Daten aus Brno.

Viele der bisher durchgeführten Studien basieren auf dem UTU-CHB Dataset, das aus 552 Fällen besteht. Ein anderer wichtiger Aspekt sind die Selektionskriterien der Daten. Bei der Planung des tschechischen Datasets wurde die Anzahl der schweren Azidosen (pH-Wert unter 7.05) auf 7 % festgelegt, was einer überdurchschnittlichen Rate entspricht (Chudáček et al. 2014). Die Quote an schweren Azidosen (pH < 7.05) in der Studie von Petrozziello beträgt 4%, ebenfalls deutlich mehr als der in epidemiologischen Studien angegebene Wert von 0.68 %.

Die Algorithmen, die sich mit Datenklassifikation beschäftigen, zeigen eine bessere Leistungsfähigkeit bei polarisierten Datenbanken. Diese Aussage konnte am Tübinger Datenset im Sinne der Steigerung des AUC-Wertes von 0.57 auf 0.6 bei Änderung der Azidoserate (< 7.05) von 0.6 % auf 3.6 % belegt werden. In der Studie der Arbeitsgruppe Ogasawara et al. wurde das Modell auf den vorselektierten "normalen" und "abnormalen" Fällen trainiert - mit einem Cut-off von 7.2 (ebenfalls 162 Fälle). Beide Arbeitsgruppen validierten ihre Modelle auf der öffentlichen Datenbank aus Brno mit einem überdurchschnittlichen Anteil pathologischer Fälle. Die Arbeitsgruppe Ogasawara et al. selektierte zusätzlich die Daten aus dieser Datenbank, sodass beide Gruppen die "normalen" und die "abnormalen" Fälle noch besser repräsentiert wurden. Wir schlagen daher vor, als Schritt zwischen dem Aufbau des KI-Modells, das Kardiotokogramme auswertet, und der Durchführung einer randomisierten prospektiven Studie eine Validierung auf einer externen Datenbank mit einem durchschnittlichen Anteil pathologischer Fälle einzuplanen. Eine solche Datenbank entspricht der klinischen Realität. Die Reimplementation des Algorithmus der Arbeitsgruppe Ogasawara auf dem Tübinger Dataset ergab ein Ergebnis von 0.6 AUC im Vergleich zu dem originalen Wert von 0.73. Vor allem wenn als endgültiges Ziel die Erstellung eines Instruments gesetzt wird, das die klinische ärztliche Tätigkeit unterstützt, und nicht der Aufbau eines Algorithmus, der polarisierte Daten sortiert.

Zahlreiche Projekte haben versucht, das Potential der künstlichen Intelligenz im Aufbau der in der Medizin gebräuchlichen Modelle zu nutzen. In der letzten Zeit wurde die Rolle des Bias in den Datenbanken und des daraus entstehenden Clever-Hans-Phänomenon zunehmend diskutiert. Eine aus der Universität Heidelberg stammende Studie untersucht den Einfluss von Hautmarkierungen in dermatoskopischen Aufnahmen auf die Effektivität eines faltenden neuronalen Netzes in der Klassifikation von Befunden in maligne und benigne. Chirurgische Markierungen der suspekten Areale erhöhten die Wahrscheinlichkeit der Einstufung des Befundes in die maligne Kategorie durch ein CNN (Winkler et al. 2019).

In unserer Studie zeigten wir die Rolle des Bias in der Konstruktion eines auf KI basierenden CTG-Auswertungssystems. Die unterschiedlichen Muster der Kardiotokogramme-eine Pause in der Ableitung während der präoperativen Vorbereitung der Patientin und der Fahrt zum OP in der Sectio-Gruppe-spielte eine Rolle in der Klassifizierung der Befunde als spontane Geburten oder Kaiserschnitte. Die beiden Gruppen zeigten signifikante Unterschiede im arteriellen Nabelschnur-pH-Wert (7.21 vs 7.27). Die pH-Vorhersage basierte unter anderem auf der Identifikation der für die Sectio-Gruppe typischen Pause in der Ableitung und der Tatsache, dass der durchschnittliche pH-Wert in dieser Gruppe signifikant höher war. Keiner dieser Faktoren hätte im Lernprozess benutzt werden sollen; es waren Faktoren, die aus medizinischer Sicht keine Rolle in der Senkung des pH-Werts spielen (präoperative Pause aus logistischen Gründen) oder nur im Tübinger Kollektiv spezifisch waren (pH-Unterschied in den Gruppen). Dieses Beispiel für Shortcut-Learning (Lernen der Eigenschaften des Datensets anstatt der echten Pathologien) zeigt die Wichtigkeit sowohl der Kenntnis und der detaillierten Analyse der eigenen Datenbank als auch des kritischen Denkens und einer Validation auf externen Daten auf.

In dieser Studie wurde die Analyse der klinischen Risikofaktoren und deren Einflusses auf den arteriellen Nabelschnur-pH-Wert untersucht. In der Literatur werden vorgeburtliche Risikofaktoren für die Senkung des arteriellen NS-pH-Wertes/Entstehung der Azidose angegeben wie: gestationales Alter, Primiparität, prolongierte Geburt (Pérez et al. 2021) und Z.n.Sectio, Z.n. IUFT, maternales Alter über 35 Jahre (Maisonneuve et al. 2011). Unsere Studie untersuchte diese und die anderen Risikofaktoren in Bezug auf die Senkung des arteriellen Nabelschnur-pH-Wertes. Die Analyse der ganzen Studienpopulation ergab

die Risikofaktoren, die mit einem erhöhten pH-Wert verbunden waren (Gestationsalter unter 37 SSW, Schätzwert unter 2,5 kg, Z.n. Sectio, Adipositas und Präeklampsie/HELLP-Syndrom). Die "protektive" Wirkung dieser Faktoren entstand aufgrund der häufigen Indikation der operativen Entbindung (54-77 % der Patientinnen aus diesen Subgruppen bekamen einen Kaiserschnitt). Die Subgruppe der spontanen Geburten schien die Korrelation zwischen den Risikofaktoren und dem Nabelschnur-pH-Wert besser abzubilden. In der Gruppe wurde ein signifikantes Risiko zur Senkung des fetalen pH-Wertes im Z.n. Sectio festgestellt und es zeigte sich im Fall der Terminüberschreitung (> 41+3 SSW, p-value 0.08) eine Tendenz dazu.

Zwei andere Faktoren in der Subpopulation der SG zeigten Signifikanz im Bezug auf die Erhöhung des Nabelschnur-pH-Wertes: ein Gestationsalter unter 37+0 SSW und ein Schätzwert unter 2500g. Die Daten zeigen, dass die pH-Werte mit steigendem Gestationsalter eine sinkende Tendenz aufweisen und diese Erscheinung physiologisch ist. Es wurde für die Interpretation der pH-Werte bei Frühgeborenen die Einführung von Perzentilkurven vorgeschlagen (Zaigham et al. 2019). 92% der Feten mit einem Schätzwert unter 2.5kg gehörten zu der Gruppe der Frühgeborenen. Ein höherer Nabelschnur-pH-Wert in den beiden Subgruppen konnte als normal angesehen werden.

Die Eingabe der Patientendaten in unser KI-Modell verbesserte in beiden Fällen die Leistungsfähigkeit des Algorithmus um 0.01 AUC. Obwohl die statistische Auswertung einen signifikanten Einfluss der ausgewählten Faktoren auf den arteriellen NS-pH-Wert beweist, war die Effizienzverbesserung des Modells niedriger als erwartet.

Für die weitere Entwicklung unseres Modells schlagen wir daher Folgendes vor:

- Erweiterung der Patientenpopulation auf fünf- bis zehntausend Fälle, um die Effektivität des Modells zu verbessern
- Automatisierung der Qualitätssicherung um die o.g. Erweiterung zu ermöglichen. Basieren des Automatisierungsprozesses auf den Ergebnissen der manuellen Qualitätssicherung

- Aufbau einer homogenen Population (spontane Geburten, reife Schwangerschaften), um das Risiko des Shortcut-Learning zu minimieren und die physiologische Gestationsalter abhängige Unterschiede auszuschließen
- Einschluss einer begrenzten Zahl der klinischen Risikofaktoren, die signifikanten Einfluss auf die Senkung des Nabelschnur-pH-Wertes zeigten (Z.n. Sectio, Terminüberschreitung, ggf. maternales Alter)

Die Patientenzahl sahen wir als eine der Hauptlimitationen der Studie. Methoden der Künstlichen Intelligenz entfalten ihr volles Potential bei der Einbindung großer Datenbanken. Idealerweise sollte die Fallzahl bei Zehntausenden oder mehr liegen. Die Automatisierung der Qualitätssicherung könnte eine Steigerung der Probandenzahl ermöglichen. Eine weitere Limitation war die Begrenzung des Outcomes zu der Geburtsazidose, was der aktuellen klinischen Praxis entspricht. Um die Interpretation der Kardiotokographie zu verbessern, sollte der Schwerpunkt nicht nur auf den Eingangsdaten (CTG, klinische Daten), sondern auch auf das fetale Outcome gelegt werden. Die Faktoren wie Notwendigkeit der postpartalen Resuscitation, Verlegung auf die neonatale Intensivstation, Dauer des Aufenthalts auf der Intensivstation, Diagnose einer ischämischen hypoxische Enzephalopathie oder neurologische Evaluation im kindlichen Alter sind ein wichtiger Bestandteil der Gesundheitsbewertung eines Neugeborenen (bez. eines Kindes) und könnten eine neue Richtung in der CTG-Auswertung setzen.

Die Erstellung eines KI-unterstützten CTG-Auswertungssystems könnte für die Entwicklung der Perinatalogie eine signifikante Bedeutung haben. Die Anwendung eines solchen Systems würde möglicherweise die Indikationsbereiche für die Kardiotokographie besser definieren. Vor allem die Patientinnengruppen, die von der CTG-Überwachung profitieren, sind von den Gruppen, die keine Dauerüberwachung brauchen, zu unterscheiden. Ein KI-Algorithmus könnte als Frühwarnsystem eingesetzt werden, das bei den Meinungsunterschieden im geburtshilflichen Team oder als Hilfe für die Berufseinsteiger im Sinne von Zweitmeinungsverfahren gelten könnte. Durch diese Einsätze könnte die weitere Senkung der fetalen und maternalen Morbidität und Mortalität erreicht werden.

## 5 Zusammenfassung

Eine der wichtigsten diagnostischen Methoden, die den ärztlichen Entscheidungsprozess in der Geburtshilfe unterstützen, ist die Kardiotokographie. Diese Überwachungsmethode besteht aus der gleichzeitigen Ableitung der fetalen Herzaktion und der uterinen Kontraktionen sowie deren Darstellung in digitaler und Papierform.

Die Sensitivität der Methode ist mit 99% hoch, die Spezifität ist jedoch mit 77% niedrig. Die Einführung der peripartalen CTG-Überwachung anstelle der intermittierenden Auskultation hat zwar zu einer Senkung der Rate an neonatalen Krampfanfällen geführt, andere Faktoren wie die perinatale Mortalität, Geburtsazidose oder Zerebralparese blieben unverändert. Seit der Verbreitung kontinuierlicher CTG-Anwendungen in der Wehenphase wurde eine Steigerung der Rate an operativen Entbindungsmodi beobachtet.

Um die Effizienz der CTG-Anwendung zu verbessern, werden Möglichkeiten der automatisierten CTG-Auswertung erforscht, unter anderem das Oxford-CTG, das STAN-CTG und das INFANT-System. Für keines dieser Systeme konnte eine bessere als die ärztliche Leistungsfähigkeit in den randomisierten Studien für das Niedrigrisiko-Kollektiv bestätigt werden.

Ziel dieser Arbeit war der Aufbau eines auf KI-Methoden basierenden CTG-Auswertungssystems. Es wurden 1394 Fälle von Einlingsschwangerschaften der Patientinnen ausgewählt, die in den Jahren 2015-2021 in der Tübinger Klinik entbunden haben. Nach manueller Qualitätssicherung sind 18,9 % der Fälle ausgeschlossen worden und die endgültige Fallzahl betrug 1131. Die wichtigsten Ausschlusskriterien waren eine Signalverlustrate über 20% und eine fragliche Qualität der arteriellen Nabelschnur-pH-Probe. Ein 30-minütiger Ausschnitt des vorgeburtlichen CTG wurde für die Analyse verwendet. Die Patientendaten enthielten die wichtigsten und am häufigsten in der Geburtshilfe vorkommenden Risikofaktoren (siehe Kapitel 2.4).

Die Patientendaten wurden statistisch in Bezug auf den arteriellen Nabelschnur-pH-Wert untersucht. In der Gesamtpopulation wirkten die Faktoren Frühgeburt, Z.n.Sectio, hypertensive SS-Erkrankung, BMI > 30 oder Schätzwert unter 2.5kg protektiv auf den arteriellen Nabelschnur-pH-Wert. Die Mehrheit dieser Patientinnen (55-77 %) erhielt einen Kaiserschnitt. In der Subpopulation der

Patientinnen, die spontan entbunden haben, zeigte der Faktor "Z.n. Sectio" eine signifikante Rolle in der Senkung des arteriellen Nabelschnur-pH-Wertes (p value 0.04) und der Faktor "Übertragung des Termins über 41+3 SSW" eine Tendenz (p-value 0.08). Diese Subpopulation bildete den Geburtsstress für den Fetus besser ab.

Es wurde ein KI-Modell erstellt, dessen Aufgabe es war, die Kardiotokogramme dem pH-Bereich zuzuordnen. Der Cut-off-Wert betrug 7.2. Zum Aufbau des Modells wurde ein CNN benutzt. Die erste Analyse ergab eine Leistungsfähigkeit des Algorithmus von 0.68 AUC. Der AUC-Zielwert in künstlicher Intelligenz ist größer als 0.7.

Bei der Konstruktion des Modells sind zwei Eigenschaften der Datenbank aufgefallen. Zum einen beinhaltete die Subgruppe der Sectio-CTG ein Muster, das einer Pause in der Ableitung bei der präoperativen Vorbereitung der Patientin entsprach. Zum anderen unterschieden sich die beiden Subgruppen signifikant, was den durchschnittlichen pH-Wert angeht (7.21 in der SG-Gruppe, 7.27 in der Sectio-Gruppe).

Die separate Analyse der Subgruppe spontaner Geburten ergab eine Leistungsfähigkeit des Algorithmus von 0.57 AUC. Dieses Phänomen wird in der Literatur als Clever-Hans-Phänomen bezeichnet, und es bedeutet, dass das Modell anhand falscher Assoziationen gelernt hat. Obwohl die statistische Analyse einen signifikanten Einfluss der Patientendaten auf den Nabelschnur-pH-Wert ergab, verbesserte die Eingabe dieser Daten in das KI-Modell dessen Leistungsfähigkeit lediglich um 0.01 AUC.

Die in dieser Arbeit erstellte Datenbank (1097 Fälle) enthält eine durchschnittliche Rate an pathologischen Fällen: 0,6 % schweren Azidosen mit einem pH-Wert unter 7.05. Nach Eingabe von weiteren 34 extrem pathologischen Fällen wuchs diese Rate auf 3.6 % und die Leistungsfähigkeit des Modells verbesserte sich auf 0.6 AUC. Die vorherigen Studien der anderen Arbeitsgruppen basierten auf Datenbanken mit einem überdurchschnittlichen Anteil an Pathologien und hatten entsprechend hohe Ergebnisse. Diese Arbeit unterstreicht die Wichtigkeit der Validation eines Modells auf einer Datenbank, die einem realen Kollektiv entspricht.

Deep Learning ist ein Wissensgebiet, das in den letzten zehn Jahren eine größte Entwicklung erlebt hat. Weitere Forschung im Thema CTG-Auswertung mit KI-Methoden sind daher wünschenswert.

## Abbildungsverzeichnis

Abbildung 1: Graphische Darstellung des Konzeptes der künstlichen Intelligenz

Abbildung 2: Schematische Darstellung der Kardiotokographie mit Positionierung der Sonden am Abdomen der Patientin (aus Barton Health)

Abbildung 3: Abbildung 3: Beispiel eines Kardiotokogramms: 1. fetale Herzaktion, 2.fetale Bewegungen, 3.uterine Kontraktionen (Hösli et al. 2017)

Abbildung 4: Abbildung 4: Ergebnisse der Qualitätssicherung: blaues Feld - gute Datenqualität. Oranges, grünes, violette und weißes Feld - Daten je nach Ausschlussgrund

Abbildung 5: Definition der fetalen Azidämie im Rahmen der Fetalblutanalyse (FBA) nach Saling

Abbildung 6: APGAR-Schema zur klinischen Beurteilung des Neugeborenen (Wert 8-10: gesundes Neugeborene, Wert 4-8: mittelschwere Adaptationsstörung, Wert < 4: schwere Asphyxie)

Abbildung 7: Darstellung der schwangeren Population nach Alter, Parität und BMI, zum Zeitpunkt der Entbindung

Abbildung 8: Maternales Risikoprofil

Abbildung 9: Aufteilung der Entbindungsmodi

Abbildung 10: Aufteilung der sekundären Kaiserschnitte nach Indikation

Abbildung 11: META - Patientendaten, CTG- Daten aus der Kardiotokogramme, BOTH - Kombination der beiden Datenkategorien

Abbildung 12: Grüne Kurve: Herzaktion des Fetus, rote Kurve: Signalverlust

Abbildung 13: Effektivität des KI-Modells auf der Subgruppe der spontanen Geburten

Abbildung 14: Daten aus den Kardiotokogrammen, BOTH - Kombination der Kardiotokogramme mit Patientendaten

Abbildung 15: Leistungsfähigkeit des Modells nach der Aufnahme von zusätzlichen pathologischen Fällen

Abbildung 16: Effizienz des externen Modells auf eigenen Daten (1131 Fälle). GlobalAveragePooling: eigenes Modell, OgasawaraNet: externes Modell

## Tabellenverzeichnis

Tabelle 1: Tabellarische Darstellung der FIGO-Kriterien. \*Repetitive Dezelerationen: treten bei mehr als 50 % aller Uteruskontraktionen auf

Tabelle 2: Mütterliches Risikoprofil - Kriterien und Erläuterungen

Tabelle 3: Fetale Charakteristika. Pc - Perzentile, PI - Pulsatilitätsindex - einer der Hauptparameter des Gefäßwiderstandes

Tabelle 4: Informationen über den peripartalen Zeitraum

Tabelle 5: Gestationsalter zum Zeitpunkt der Geburt, fetale Biometrie und Farbdoppler aus der Aufnahmeuntersuchung-maximal 2 Wochen vor der Entbindung. M-Mittelwert, SA-Standardabweichung, PI-Pulsatility Index, Pc-Perzentile, AFI-Amniotic Fluid Index, SD-Single Depot

Tabelle 6: Charakteristika der Outcome-Daten. A-primäre Studienpopulation (1097 Fälle), B-zusätzliche Fälle mit kindlichem arteriellem pH-Wert < 7,0 (34 Fälle)

Tabelle 7: Analyse der Patientendaten in Bezug auf arterielle Nabelschnur-pH-Werte in der Gesamtpopulation (1097 Fälle). P-value unter 0.05 weist auf statistische Signifikanz hin

Tabelle 8: Analyse der Patientendaten in Bezug auf arterielle Nabelschnur-pH-Werte in der Subpopulation der spontanen Geburten. P-value unter 0.05 weist auf statistische Signifikanz hin. \*p-value entspricht dem Wert aus der Gesamtpopulation (siehe Tabelle 7)

## Literaturverzeichnis

Abou-Dakn, Michael. "Vaginale Geburt Am Termin - AWMF." *AWMF Leitlinienregister*, 2020,

[https://www.awmf.org/uploads/tx\\_szleitlinien/015-083m\\_S3\\_Vaginale-Geburt-am-Termin\\_2021-01\\_1.pdf](https://www.awmf.org/uploads/tx_szleitlinien/015-083m_S3_Vaginale-Geburt-am-Termin_2021-01_1.pdf). Accessed 24 Jan. 2023.

Alfirevic, Zarko, et al. "Continuous Cardiotocography (CTG) as a Form of Electronic Fetal Monitoring (EFM) for Fetal Assessment during Labour." *Cochrane Database of Systematic Reviews*, 2006, <https://doi.org/10.1002/14651858.cd006066>.

Anceschi, M. M., et al. "Validity of Short Term Variation (STV) in Detection of Fetal Acidemia." *Journal of Perinatal Medicine*, vol. 31, no. 3, 2003, <https://doi.org/10.1515/jpm.2003.031>.

Ayres-de-Campos, Diogo, et al. "Figo Consensus Guidelines on Intrapartum Fetal Monitoring: Cardiotocography." *International Journal of Gynecology & Obstetrics*, vol. 131, no. 1, 2015, pp. 13–24., <https://doi.org/10.1016/j.ijgo.2015.06.020>.

Barton Health. "External and Internal Heart Rate Monitoring of the Fetus." *Barton Health | Health Library | Staywell Health Library | External and Internal Heart Rate Monitoring of the Fetus*, [https://www.bartonhealth.org/tahoe/health-library/external-and-internal-heart-rate-monitoring-of-the-20585.aspx?iid=92\\_P07776](https://www.bartonhealth.org/tahoe/health-library/external-and-internal-heart-rate-monitoring-of-the-20585.aspx?iid=92_P07776). Accessed 15 Feb. 2023.

Becker, Jeroen H., et al. "St Analysis of the Fetal Electrocardiogram in Intrapartum Fetal Monitoring: A Meta-Analysis." *Obstetrics & Gynecology*, vol. 119, no. 1, Jan. 2012, pp. 145–154., <https://doi.org/10.1097/AOG.0b013e31823d8230>.

Bender, H.G., and K. Vetter. "40 Jahre Kardiotokographie." *Der Gynäkologe*, vol. 42, no. 5, 2009, pp. 327–327., <https://doi.org/10.1007/s00129-009-2357-3>.

Bilardo, C. M., et al. "Severe Fetal Growth Restriction at 26-32 Weeks: Key Messages from the Truffle Study." *Ultrasound in Obstetrics & Gynecology*, vol. 50, no. 3, 2017, pp. 285–290., <https://doi.org/10.1002/uog.18815>.

Bobrow, C. S, and P. W Soothill. "Causes and Consequences of Fetal Acidosis." *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 80, no. 3, 1999, <https://doi.org/10.1136/fn.80.3.f246>.

Brocklehurst, Peter, et al. "Computerised Interpretation of Fetal Heart Rate during Labour (Infant): A Randomised Controlled Trial." *The Lancet*, vol. 389, no. 10080, 2017, pp. 1719–1729., [https://doi.org/10.1016/s0140-6736\(17\)30568-8](https://doi.org/10.1016/s0140-6736(17)30568-8).

Chudáček, Václav, et al. "Open Access Intrapartum CTG Database." *BMC Pregnancy and Childbirth*, vol. 14, no. 1, 2014, <https://doi.org/10.1186/1471-2393-14-16>.

Copeland, B.J. "Artificial Intelligence." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., 2022, <https://www.britannica.com/technology/artificial-intelligence>. Accessed 24 Jan. 2023.

Dongus, Kathrin Verena. "Klinische Machbarkeitsstudie: Messung Der Maternalen Herzfrequenz Durch Einen in Den Tokografie-Transducer Integrierten Infrarot-Sensor Im Rahmen Der Kardiotokografie." *Universität Tübingen*, Universitätsbibliothek, 2011, pp. 17–26, [https://publikationen.uni-tuebingen.de/xmlui/bitstream/handle/10900/45837/pdf/Diss\\_Abgabe.pdf?sequence=1&isAllowed=y](https://publikationen.uni-tuebingen.de/xmlui/bitstream/handle/10900/45837/pdf/Diss_Abgabe.pdf?sequence=1&isAllowed=y). Accessed 24 Jan. 2023.

Dunn, P M. “Adolphe Pinard (1844-1934) of Paris and Intrauterine Paediatric Care.” *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 91, no. 3, 2005, <https://doi.org/10.1136/adc.2005.074518>.

Goeschen, K. “Interpretation Und Konsequenzen Der Kardiotokographie (CTG).” *Gynäkologie Und Geburtshilfe 1990*, 1991, pp. 609–614., [https://doi.org/10.1007/978-3-642-76124-9\\_343](https://doi.org/10.1007/978-3-642-76124-9_343).

Götz, Rudolf, et al. *Ultraschalldiagnostik – Geräteaufbau, Doppler Und Methoden*. Springer, Berlin, Heidelberg, 2015.

Hernández-Orallo, José. “Gazing into Clever Hans Machines.” *Nature Machine Intelligence*, vol. 1, no. 4, 2019, pp. 172–173., <https://doi.org/10.1038/s42256-019-0032-5>.

Hruban, Lukáš, et al. “Agreement on Intrapartum Cardiotocogram Recordings between Expert Obstetricians.” *Journal of Evaluation in Clinical Practice*, vol. 21, no. 4, 2015, pp. 694–702., <https://doi.org/10.1111/jep.12368>.

Hösli, Irene, and Johanna Büchel. “Intrapartales Kardiotokogramm.” *Der Gynäkologe*, vol. 50, no. 9, 2017, pp. 695–709., <https://doi.org/10.1007/s00129-017-4114-3>.

Jiménez, Marta. “Unboxing the Black Box of Ai.” *Utrecht University*, 14 Jan. 2022, <https://www.uu.nl/en/organisation/in-depth/unboxing-the-black-box-of-ai>.

Accessed 06 Feb. 2023

Jonsson, M, et al. “Metabolic Acidosis at Birth and Suboptimal Care - Illustration of the Gap between Knowledge and Clinical Practice.” *BJOG: An International Journal*

*of Obstetrics & Gynaecology*, vol. 116, no. 11, 2009, pp. 1453–1460.,  
<https://doi.org/10.1111/j.1471-0528.2009.02269.x>.

Lange, Shannon, et al. “National, Regional, and Global Prevalence of Smoking during Pregnancy in the General Population: A Systematic Review and Meta-Analysis.” *The Lancet Global Health*, vol. 6, no. 7, 2018,  
[https://doi.org/10.1016/s2214-109x\(18\)30223-7](https://doi.org/10.1016/s2214-109x(18)30223-7).

Lasch, Lidia, and Sabine Fillenberg. “15.4 Schwangerenvorsorge.” *Basiswissen Gynäkologie Und Geburtshilfe*, Springer, Berlin, 2017, pp. 184–196.

Maisonneuve, Emeline, et al. “Risk Factors for Severe Neonatal Acidosis.” *Obstetrics & Gynecology*, vol. 118, no. 4, 2011, pp. 818–823.,  
<https://doi.org/10.1097/aog.0b013e31822c9198>.

Nahm, Francis Sahngun. “Receiver Operating Characteristic Curve: Overview and Practical Use for Clinicians.” *Korean Journal of Anesthesiology*, vol. 75, no. 1, 2022, pp. 25–36., <https://doi.org/10.4097/kja.21209>.

Nelson, Karin B., et al. “Uncertain Value of Electronic Fetal Monitoring in Predicting Cerebral Palsy.” *New England Journal of Medicine*, vol. 334, no. 10, 1996, pp. 613–619., <https://doi.org/10.1056/nejm199603073341001>.

Ogasawara, Jun, et al. “Deep Neural Network-Based Classification of Cardiotocograms Outperformed Conventional Algorithms.” *Scientific Reports*, vol. 11, no. 1, 2021, <https://doi.org/10.1038/s41598-021-92805-9>.

O'Sullivan, M., et al. “Classification of Fetal Compromise during Labour: Signal Processing and Feature Engineering of the Cardiotocograph.” *2021 29th European*

*Signal Processing Conference (EUSIPCO)*, 2021b,  
<https://doi.org/10.23919/eusipco54536.2021.9616289>.

O’Sullivan, M. E., et al. “Challenges of Developing Robust AI for Intrapartum Fetal Heart Rate Monitoring.” *Frontiers in Artificial Intelligence*, vol. 4, 2021a,  
<https://doi.org/10.3389/frai.2021.765210>.

Pardey, James, et al. “A Computer System for the Numerical Analysis of Nonstress Tests.” *American Journal of Obstetrics and Gynecology*, vol. 186, no. 5, 2002, pp. 1095–1103., <https://doi.org/10.1067/mob.2002.122447>.

“Perinatalsterblichkeit in Deutschland, West- Und Ostdeutschland (1955-2020).” *Bundesinstitut Für Bevölkerungsforschung*,  
<https://www.bib.bund.de/DE/Fakten/Fakt/S44-Perinatalsterblichkeit-West-Ost-ab-1955.html>.

Petrozziello, Alessio, et al. “Multimodal Convolutional Neural Networks to Detect Fetal Compromise during Labor and Delivery.” *IEEE Access*, vol. 7, 2019, pp. 112026–112036., <https://doi.org/10.1109/access.2019.2933368>.

Pérez, María Luisa, et al. “Analysis of Factors Associated with Variability and Acidosis of the Umbilical Artery Ph at Birth.” *Frontiers in Pediatrics*, vol. 9, 2021,  
<https://doi.org/10.3389/fped.2021.650555>.

Pérez, María Luisa, et al. “Analysis of Factors Associated with Variability and Acidosis of the Umbilical Artery Ph at Birth.” *Frontiers in Pediatrics*, vol. 9, 2021,  
<https://doi.org/10.3389/fped.2021.650555>.

Spencer, John A., et al. "The Intrapartum CTG Prior to Neonatal Encephalopathy at Term: A Case-Control Study." *BJOG: An International Journal of Obstetrics and Gynaecology*, vol. 104, no. 1, 1997, pp. 25–28.,  
<https://doi.org/10.1111/j.1471-0528.1997.tb10643.x>.

"Recommendations: Intrapartum Care for Healthy Women and Babies: Guidance." *NICE*, Dec. 2022,  
<https://www.nice.org.uk/guidance/cg190/chapter/Recommendations#monitoring-during-labour>.

Ross, Michael G., and Rageev Gala. "Use of Umbilical Artery Base Excess: Algorithm for the Timing of Hypoxic Injury." *American Journal of Obstetrics and Gynecology*, vol. 187, no. 1, 2002, pp. 1–9.,  
<https://doi.org/10.1067/mob.2002.123204>.

Schneider, Henning, et al. *Die Geburtshilfe*. Springer, 2016.

Vintzileos, A. "Intrapartum Electronic Fetal Heart Rate Monitoring versus Intermittent Auscultation: A Meta-Analysis." *Obstetrics & Gynecology*, vol. 85, no. 1, 1995, pp. 149–155., [https://doi.org/10.1016/0029-7844\(94\)00320-d](https://doi.org/10.1016/0029-7844(94)00320-d).

Warrick, Philip A., et al. "A Machine Learning Approach to the Detection of Fetal Hypoxia during Labor and Delivery." *AI Magazine*, vol. 33, no. 2, 2012, p. 79.,  
<https://doi.org/10.1609/aimag.v33i2.2412>.

Winkler, Julia K., et al. "Association between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition." *JAMA Dermatology*, vol. 155, no. 10, 2019, p. 1135., <https://doi.org/10.1001/jamadermatol.2019.1735>.

Zaigham, Mehreen, et al. "Gestational Age-Related Reference Values for Apgar Score and Umbilical Cord Arterial and Venous Ph in Preterm and Term Newborns." *Acta Obstetricia Et Gynecologica Scandinavica*, vol. 98, no. 12, 2019, pp. 1618–1623., <https://doi.org/10.1111/aogs.13689>.

## **Erklärung zum Eigenanteil der Dissertationsschrift**

Die Arbeit wurde in der Universitäts-Frauenklinik Tübingen unter Betreuung von Professor Dr. med. H. Abele durchgeführt.

Die statistische Auswertung erfolgte eigenständig durch mich. Die Methoden der Künstlichen Intelligenz wurden von Herrn J.Vetter, PhD-Student (Machine Learning for Science, Informatik, Universität Tübingen), entwickelt.

Ich versichere, das Manuskript selbständig verfasst zu haben und keine weiteren als die von mir angegebenen Quellen verwendet zu haben.

Agata Wolny

Tübingen, den 26.08.2023

## **Danksagung**

Mein besonderer Dank gilt Herrn Professor Dr. med. H. Abele, dem Leiter der Abteilung für Perinatalmedizin, für die kompetente Betreuung und Motivation. Ohne seine Professionalität und seine Offenheit für neue Ideen wäre diese Arbeit nie zustande gekommen.

Herzlich danken möchte ich auch Herrn Dr. med. P. Jakubowski, Oberarzt der Geburtshilfe, für die vielen Ratschläge.

Des Weiteren danke ich allen Mitarbeitern der CTG-Arbeitsgruppe, vor allem Herrn Professor Dr.-Ing. O. Kohlbacher, Herrn Professor Dr. rer. nat. Jakob Macke, Herrn Dr. Tjeerd Dijkstra, Herrn Dipl.-Ing. Jörg Marstaller und Herrn J.Vetter, PhD-Student in Machine Learning for Science (Fachbereich Informatik, Universität Tübingen), für die gute und angenehme interdisziplinäre Mitarbeit.