

Научная статья  
УДК 618.5-07

**ТАЕВА**  
**Ульяна Евгеньевна**

специалитет, Петрозаводский государственный  
институт (Петрозаводск, Россия),  
[taevaullyana@gmail.com](mailto:taevaullyana@gmail.com)

## **ПРОБЛЕМА СУБЪЕКТИВНОСТИ ВИЗУАЛЬНОЙ ИНТЕРПРЕТАЦИИ КАРДИОТОКОГРАФИИ И ПЕРСПЕКТИВЫ ЕЁ АВТОМАТИЗАЦИИ**

### **Научный руководитель:**

Ившин Александр  
Анатольевич

### **Рецензент:**

Болдина Юлия Сергеевна

Статья поступила: 07.05.2026;

Принята к публикации: 27.06.2026;

Размещена в сети: 27.06.2026.

**Аннотация.** Кардиотокография (КТГ) – основной метод интранатального мониторинга плода, однако её визуальная интерпретация субъективна: межэкспертное согласие при оценке не превышает 48 %, и ни одно руководство не обеспечивает одновременно высокой чувствительности и специфичности. Обзор 67 публикаций за 2015–2025 годы показал, что алгоритмы МО превосходят визуальную оценку по воспроизводимости, достигая точности 87–99 %, но применимость ограничена несбалансированностью обучающих данных и снижением точности на новых выборках. Перспективное направление – интеграция ИИ в системы поддержки врачебных решений.

**Ключевые слова:** кардиотокография, машинное обучение, межэкспертная вариабельность, система поддержки принятия врачебных решений, интерпретация КТГ, мониторинг плода, искусственный интеллект

**Благодарности.** Исследования, описанные в данной работе, были проведены в рамках проекта «Разработка программно-аппаратного комплекса, интегрирующего модель машинного обучения для распознавания паттернов кардиотокографии (КТГ) с веб-сервисом для загрузки данных», поддержанного в рамках Программы поддержки НИОКР студентов, аспирантов и лиц, имеющих ученую степень, обеспечивающих значительный вклад в инновационное развитие отраслей экономики и социальной сферы Республики Карелия, в 2026 году, финансируемой Правительством Республики Карелия (Договор №1-Г25 от 26.12.2025 между ФГБОУ ВО «Петрозаводский государственный университет» и Фондом венчурных инвестиций Республики Карелия).

**Для цитирования:** Таева У. Е. Проблема субъективности визуальной интерпретации кардиотокографии и перспективы её автоматизации // StudArctic Forum. 2026. Т. 11, № 2. С. 42–55.

Кардиотокография (КТГ) – метод непрерывной регистрации частоты сердечных сокращений (ЧСС) плода и сократительной активности матки посредством ультразвукового датчика, размещённого на передней брюшной стенке беременной. КТГ применяется в клинической практике с конца 1960-х годов и по сей день остаётся основным методом интранатального мониторинга состояния плода в современном акушерстве. Метод неинвазивен, относительно недорог и повсеместно распространён. Так, по имеющимся данным, в США КТГ используется при ведении 89 % одноплодных беременностей в родах [Mendis, 2025: 123].

Несмотря на многолетнее применение, визуальная интерпретация кардиотокограмм остаётся субъективным процессом. По данным Lukáš Hruban и соавт., общая доля согласованности между девятью опытными акушерами при анализе 634 интранатальных КТГ составила лишь 48 %, при этом для патологического класса – 41 % [Hruban: 694]. Аналогичная картина наблюдается и в крупном международном онлайн-исследовании Imape Ven M'Varek и соавт., где 120 специалистов по плёнкам КТГ правильно предсказали уровень рН плода менее 7,15 лишь в 58 % случаев, а чувствительность составила 0,58 при специфичности 0,63 [Ven M'Varek: 136]. Субъективность интерпретации КТГ напрямую влияет на клинические решения: Vejux и соавт. показали, что 70 % случаев кесарева сечения по поводу «тревожного» КТГ при слепой ретроспективной оценке расценивались экспертами как необоснованные [Vejux: 1816]. Всё это ставит закономерный вопрос: может ли формализованный, воспроизводимый алгоритм обеспечить более стабильную и объективную оценку КТГ там, где человеческая экспертиза даёт сбой?

За последнее десятилетие накоплен значительный массив исследований, посвящённых применению различных алгоритмов машинного обучения (МО) для автоматизированного анализа КТГ, начиная с классических методов (метод опорных векторов, случайного леса, деревьев решений) и заканчивая свёрточными нейронными сетями и трансформерными архитектурами. Есть основания полагать, что именно эти подходы способны устранить субъективный фактор и тем самым повысить точность и согласованность диагностики состояния плода. Вместе с тем вопрос о том, при каких условиях и с какими ограничениями это реализуемо на практике, остаётся дискуссионным.

Данная работа представляет собой систематический обзор литературы за период 2015–2025 годов, охватывающий исследования в двух взаимосвязанных областях: изучения межэкспертной вариабельности при интерпретации КТГ и разработки методов её автоматизированного анализа. Цель обзора – систематизировать имеющиеся данные о методах интерпретации КТГ, выявить ключевые ограничения современных подходов и оценить перспективы применения систем поддержки врачебных решений на основе искусственного интеллекта. Для достижения цели поставлены следующие задачи:

- 1) охарактеризовать масштаб межэкспертной вариабельности при визуальной интерпретации КТГ;
- 2) сравнить диагностические возможности действующих клинических руководств;
- 3) систематизировать методы МО для автоматизированного анализа КТГ;
- 4) выявить ключевые ограничения существующих подходов с точки зрения клинической применимости;
- 5) оценить перспективы интеграции интеллектуальных систем поддержки врачебных решений в реальную клиническую практику мониторинга состояния плода.

Поиск литературы проводился в следующих электронных базах данных: PubMed, IEEE Xplore, DOAJ (Directory of Open Access Journals), eLIBRARY.RU и MDPI. Применялись поисковые запросы, включавшие термины «cardiotocography», «CTG», «fetal heart rate», «machine learning», «deep learning», «interobserver variability», «automated analysis» и их комбинации. Поиск охватывал публикации за период с 1 января 2015 по 31 декабря 2025 года на английском и русском языках.

В обзор включались: (1) исследования диагностических тестов, проспективные и ретроспективные когортные исследования, рандомизированные контролируемые испытания, исследования «случай-контроль» и технические валидации систем интерпретации КТГ, содержащие числовые данные о производительности; (2) исследования, охватывающие беременных с гестационным сроком  $\geq 32$  недель при одноплодной беременности; (3) работы, посвящённые как визуальной интерпретации КТГ по клиническим руководствам (FIGO-2015, NICE-2022, SWE и другим), так и алгоритмам автоматической интерпретации, включая

коммерческие системы и модели МО/глубокого обучения; (4) публикации, представляющие хотя бы один из следующих результатов: показатели диагностической точности (чувствительность, специфичность, AUC-ROC), классификацию КТГ или оценку согласованности (коэффициент Каппа, внутриклассовый коэффициент корреляции).

Из обзора исключались: обзорные статьи, письма в редакцию, тезисы без полного текста, протоколы без результатов, исследования только на животных, работы, использующие исключительно синтетические данные КТГ без валидации на реальных записях, а также статьи, не предоставляющие количественных данных о результатах.

Первичный поиск выявил 497 публикаций. После удаления 207 дубликатов с использованием платформы Rayyan AI и менеджера ссылок Zotero 290 статей прошли скрининг по заголовкам и аннотациям. По результатам этого этапа 120 публикаций были отобраны для полнотекстового анализа. На финальном этапе скрининга по полным текстам с учётом критериев включения/исключения в обзор включены 67 исследований.

Включённые работы опубликованы в период 2015–2025 годов и представлены исследованиями из Европы, Азии, Северной Америки, Африки и Австралии. По тематике публикации распределились на два основных блока: исследования межэкспертной вариабельности и сравнения клинических руководств и разработки алгоритмов МО/глубокого обучения для анализа КТГ. Ряд работ охватывает оба направления, исследуя автоматизацию именно как средство преодоления субъективности.

\* \* \* \* \*

Межэкспертная вариабельность при интерпретации КТГ – это давно известное и хорошо задокументированное явление, не связанное с недостаточной квалификацией конкретных специалистов. Оно воспроизводится в исследованиях из разных стран, с разными участниками и по разным клиническим руководствам. Наиболее масштабное подтверждение этому – чешское исследование Hruban и соавт., в котором девять опытных акушеров из шести клинических центров независимо оценили 634 интранатальных КТГ: общая доля согласованных оценок составила лишь 48 %, а чувствительность прогнозирования ацидоза при консенсусном голосовании не превысила 27 % [Hruban: 694]. Принципиально важно, что ни уровень опыта специалистов, ни место их работы на этот результат не повлияли. В крупном международном онлайн-исследовании Imane Ben M'Barek и соавт. 120 акушерских специалистов из нескольких стран правильно предсказали уровень pH плода ниже 7,15 лишь в 58 % случаев, причём наибольшие трудности возникали в «серой зоне» при pH между 7,05 и 7,20, где правильный ответ давался лишь в 48 % случаев. Ни профессиональная группа, ни стаж работы на точность не влияли [Ben M'Barek: 136]. Amadori и соавт. подтвердили ту же закономерность: при анализе 73 КТГ восьмью операторами по FIGO-2015 итоговая категоризация записи как «подозрительной» или «патологической» не достигала согласия между участниками [Amadori: 2753]. Таким образом, низкая воспроизводимость интерпретации КТГ – не следствие недостаточной подготовки, а системное свойство, характерное для визуального анализа в целом.

Вместе с тем вариабельность распределена неравномерно в зависимости от конкретного оцениваемого параметра. Клинический анализ КТГ традиционно включает оценку нескольких ключевых характеристик: базальной частоты сердечных сокращений плода, вариабельности, акцелераций (кратковременных учащений ритма, свидетельствующих о реактивности плода) и децелераций, то есть урежений ЧСС. Именно на основании совокупности этих параметров и принимается решение, к какой категории относится плёнка КТГ, к нормальной, подозрительной или патологической. Linas Rovas и соавт. (2025) установили, что базальный ритм, вариабельность и наличие децелераций оценивались с умеренным или хорошим согласием ( $\kappa=0,47-0,80$ ), тогда как итоговая категоризация давала

лишь плохое или умеренное согласие ( $\kappa=0,19-0,44$ ) [Rovas: 55]. Mariana Rei и соавт. (2016) зафиксировали аналогичный паттерн. Наличие децелераций оценивалось с высокой долей согласия ( $P_a=0,92$ ), тогда как их тип (поздние или переменные) значительно хуже ( $\kappa=0,13-0,25$ ) [Rei: 27]. Serena Neri и соавт. в двух работах (2024, 2025) подтвердили, что наименее воспроизводимым параметром остаётся вариабельность ( $\kappa = 0,20$ ) [Neri, 2024: 1]; [Neri, 2025: 1]. Следовательно, проблема заключается не в КТГ-методе как таковом, а в сложности финального синтетического суждения, то есть именно того шага, на котором принимается клиническое решение, и который труднее всего формализовать.

Субъективность интерпретации КТГ влечёт прямые клинические и правовые последствия. Nadege Vejuh и соавт. (2017) показали, что 70 % кесаревых сечений, выполненных по поводу «тревожного» КТГ, при слепой ретроспективной оценке расценивались экспертами как необоснованные; при этом две трети таких операций проводились ночью, что можно расценивать как косвенное указание на роль усталости персонала [Vejuh: 1816]. Это наглядный пример гипердиагностики. Но существует и обратная сторона: S Kundu и соавт. (2017) предложили семи экзаменаторам с разным опытом классифицировать 300 КТГ по FIGO и одновременно предсказать рН пуповинной артерии плода. Корреляция между классификацией КТГ и реальным рН оказалась близка к нулю для всех экзаменаторов ( $R^2=0,06-0,17$ ), а доля пропущенных случаев ацидоза составила от 20 % у наиболее опытного специалиста до 70 % у наименее опытного [Kundu: 897]. Примечательно, что субъективность не исчезает даже в таком ответственном контексте как судебный. Laura Sabiani и соавт. (2015) изучили согласованность 22 французских акушеров, выступающих экспертами в судах по делам о врачебных ошибках. Межэкспертное согласие при классификации патологических КТГ оказалось крайне низким ( $\kappa=0,11-0,18$ ), а знание о неблагоприятном исходе значимо меняло суждения. Эксперты чаще оправдывали принятые решения ретроспективно, что является классическим смещением [Sabiani: 856]. В совокупности эти данные свидетельствуют о том, что субъективность интерпретации КТГ несёт измеримые клинические и юридические издержки, которые не могут быть устранены одной лишь стандартизацией процедур.

Часть описанной вариабельности объясняется не только субъективностью восприятия, но и принципиальными различиями между самими клиническими руководствами, а именно их терминологией, пороговыми значениями и критериями классификации. Этому вопросу посвящён следующий раздел.

Одним из очевидных ответов на проблему субъективности стала разработка стандартных клинических руководств по интерпретации КТГ. В настоящее время в международной практике используется несколько конкурирующих систем: трёхуровневые FIGO-2015, NICE (в редакциях 2017 и 2022), ACOG и NICHD, пятиуровневые CNGOF-2013 и система Parer-Ikeda, а также национальные шведские шаблоны (SWE-09 и SWE-17). Предполагалось, что формализация критериев интерпретации снизит субъективный фактор. Данные включённых исследований свидетельствуют о том, что этого не произошло, и причина кроется не в качестве отдельных руководств, а в природе самой задачи.

Наиболее прямое сравнение диагностической точности провели Clara Zamora del Pozo и соавт. (2021): три независимых рецензента классифицировали 150 КТГ по четырём системам – FIGO, ACOG, NICE и Chandraran. AUC для предсказания ацидемии новорождённых составил около 0,66 для всех четырёх систем без значимых различий между ними [Zamora del Pozo: 8498]. Denise Kling и соавт. (2024) выявили фундаментальный компромисс, присущий всем пяти изученным шаблонам без исключения: системы с высокой чувствительностью (SWE-09 и NICE-22, по 92 %) неизбежно давали низкую специфичность (~62 %), тогда как высокоспецифичные системы (FIGO-15 и SWE-17, ~90–91 %) имели

крайне низкую чувствительность (39–42 %) [Kling: 1]. Иными словами, выбор руководства определяет не столько точность диагностики, сколько то, какой тип ошибки считается менее приемлемым: пропустить патологию или избыточно вмешаться.

Попытки преодолеть это ограничение за счёт создания более детализированных руководств с большим количеством учитываемых переменных также не дали ожидаемого результата. Большинство существующих систем используют трёхуровневую классификацию КТГ – нормальная, подозрительная и патологическая. Однако ряд руководств предлагает более детальное деление. Например, пятиуровневая система CNGOF-2013 выделяет категории от «нормальной» до «претерминальной», а система Parer–Ikeda кодирует риск ацидемии по цветовой шкале от зелёного до красного. Предполагалось, что большее число градаций позволит точнее разграничить клинические ситуации и снизить субъективность. Charles Garabedian и соавт. (2017) напрямую сравнили двухуровневую систему, NICHD-2008, пятиуровневую CNGOF-2013 и трёхуровневую FIGO-2015 на 100 КТГ: межэкспертное согласие оказалось умеренным для каждой из них ( $\alpha$  Krippendorff'a от 0,48 до 0,59), причём пятиуровневая CNGOF не превзошла более простые системы [Garabedian: 131]. Sabina Martí Gamboa и соавт. (2017), сравнив трёхуровневую FIGO и пятиуровневую систему Parer на 202 КТГ-записях, зафиксировали ту же закономерность: общее межэкспертное согласие для обеих систем оказалось практически идентичным ( $\kappa=0,460$  и  $0,466$  соответственно), при этом максимальное расхождение между экспертами наблюдалось именно в промежуточных категориях, то есть там, где клиническая неопределённость наиболее высока и цена ошибки наиболее значима [Martí Gamboa: 508].

Если все системы одинаково несовершенны по диагностической точности и согласованности, можно было бы ожидать, что их выбор не имеет принципиального значения. Однако данные показывают обратное: разные руководства по-разному формируют клинические решения на одном и том же материале. Meena Bhatia и соавт. (2017) установили, что NICE-2014 распознавал патологические КТГ лишь в 47,6 % случаев, тогда как NICE-2007 – в 91,0 %, а FIGO-2015 – в 76,2 % [Bhatia: 89]. Разница в 43 процентных пункта между двумя версиями одного и того же руководства означает, что пациент с одним и тем же КТГ получит или не получит вмешательство в зависимости от того, какой стандарт применяет конкретное учреждение. Nadege Vejuh и соавт. (2017) показали, что это влияние распространяется и на воспроизводимость решений: при анализе 100 кесаревых сечений по поводу «тревожного» КТГ межэкспертное согласие при оценке обоснованности операции было значимо выше при использовании FIGO, чем CNGOF ( $\kappa=0,331$  против  $0,209$ ;  $p<0,001$ ) [Vejuh: 1816].

Таким образом, существующие клинические руководства оказываются перед нерешённым противоречием: ни одно из действующих руководств не способно одновременно обеспечить высокую диагностическую точность, удовлетворительную межэкспертную согласованность и стабильность клинических решений. Именно это противоречие создаёт объективную потребность в подходах, выходящих за рамки визуальной интерпретации – алгоритмических системах, применяющих одни и те же правила последовательно, без вариаций и независимо от усталости или опыта конкретного специалиста.

Параллельно с изучением межэкспертной вариабельности за последнее десятилетие сформировалось обширное направление исследований, посвящённых автоматизированному анализу КТГ. Логично предположить, что если оценка одной и той же записи разными специалистами даёт разные результаты, то алгоритм, применяющий одни и те же правила последовательно и без вариаций, должен справиться лучше. Проверка этой идеи на практике выявила как реальные достижения, так и серьёзные ограничения, которые пока удерживают большинство разработок на уровне исследовательских прототипов.

Классические алгоритмы машинного обучения, такие как метод опорных векторов, случайный лес, деревья решений и их комбинации устойчиво демонстрируют высокую точность при классификации КТГ. Vinayaka Nagendra и соавт. (2017) получили точность выше 96 % при сравнении этих методов в системе мониторинга реального времени [Nagendra: 1]. Astik Kumar Pradhan и соавт. (2021) достигли точности 0,99 для случайного леса [Pradhan: 239]. Объединение нескольких классификаторов в ансамбли, когда итоговое решение принимается голосованием или взвешиванием нескольких моделей, последовательно улучшает результаты по сравнению с одиночными алгоритмами, например, K Aditya Shastry и соавт. (2024) достигли лучших показателей среди базовых классификаторов именно за счёт адаптивного взвешивания [Shastry: 1], а Junyuan Feng и соавт. (2023) показали точность 0,95 на публичных и 0,92 на частных данных, объединив метод опорных векторов, XGBoost и случайный лес [Feng: 273].

Однако за этими цифрами скрывается принципиальная проблема. Подавляющее большинство исследований опирается на один и тот же публичный набор данных, а именно UCI CTG Dataset, содержащий 2126 записей КТГ, размеченных тремя экспертами по трём классам: нормальный, подозрительный и патологический. Эта база данных несбалансирована, и нормальные случаи в ней составляют около 77 % всех записей, а патологические лишь около 9 %. Это серьёзная проблема для машинного обучения, так как модель, обученная на таких данных, научится почти всегда предсказывать «норму» и при этом будет показывать высокую точность просто потому, что нормальных случаев большинство. При этом именно патологические случаи, которых меньшинство, являются клинически наиболее значимыми. Irem Nazli и соавт. (2025) специально исследовали этот эффект. Когда они применили метод искусственного увеличения числа редких случаев (SMOTE) для выравнивания классов, ранжирование алгоритмов существенно изменилось, а LightGBM показал сбалансированную точность 91,34 %, что честнее отражает реальную диагностическую ценность [Nazli: 1250]. Схожую картину зафиксировали Muhammad Anwar Ma'sum и соавт. (2019): балансировка данных критически влияла на производительность глубоких нейросетей [Ma'sum: 51]. Второй широко используемый датасет – STU-UHV, содержащий 552 интранатальных КТГ-записи из Чешского технического университета с верифицированными данными о рН пуповинной артерии. Он меньше UCI по объёму, но более однороден клинически и используется в работах, где целью является не классификация по экспертным меткам, а предсказание реального исхода [Francis, 2022: 1]; [Cömert, 2017: 1].

Методы глубокого обучения, прежде всего свёрточные нейронные сети, которые умеют самостоятельно находить закономерности в сигнале без ручного выделения признаков, открывают принципиально иные возможности. Alessio Petrozziello и соавт. (2018, 2019) из Оксфордского университета первыми систематически применили их к крупному клиническому датасету из более чем 35 000 родов. Чувствительность CNN составила 42 % против 30 % при визуальном анализе при сопоставимых ложноположительных ставках, а на двух внешних датасетах модель также превзошла методы с ручным выделением признаков [Petrozziello, 2018: 5866]; [Petrozziello, 2019: 112026]. Jun Ogasawara и соавт. (2021) разработали архитектуру, одновременно анализирующую сигнал частоты сердечных сокращений плода и сигнал маточных сокращений, обучив её предсказывать неонатальные исходы напрямую, минуя экспертные суждения. AUC в этом исследовании составил 0,73 на японском клиническом датасете, что статистически значимо превзошло рекуррентные нейросети [Ogasawara: 13367]. Yared Daniel Daydulo и соавт. (2022) преобразовали сигнал КТГ в визуальное изображение, отражающее его частотные характеристики во времени, и затем применили нейросеть ResNet, заранее обученную на миллионах изображений. Такой

подход (перенос знаний) позволил достичь точности 98,7 % на STU-UHB [Daydulo: 329]. Daniel Asfaw и соавт. (2023) обучили комбинированную модель на 51 449 родах из Оксфорда, параллельно обрабатывая временной ряд ЧСС и его визуальное представление [Asfaw: 730]. Новейшие архитектуры – трансформеры, изначально разработанные для обработки текста и умеющие улавливать долгосрочные зависимости в последовательностях (PatchCTG, Khan 2025), и сети Колмогорова–Арнольда, предлагающие более гибкое математическое описание нелинейных зависимостей (Wong 2025), расширяют арсенал подходов к анализу КТГ [Khan: 1]; [Wong: 345].

Рост сложности архитектур обострил проблему, критически важную для клинического применения: интерпретируемость. Когда система выдаёт заключение «патологическое КТГ», врач должен понимать, на каком основании это сделано, иначе он не может доверять системе. Ответом на этот вызов стало применение инструментов объяснимого ИИ, то есть методов, позволяющих количественно оценить вклад каждого признака сигнала в итоговое решение модели. Junyuan Feng и соавт. (2023) показали, что ключевыми признаками для их модели оказались акцелерации и доля времени с аномальной краткосрочной вариабельностью. Это результат, совпадающий с тем, что клиницисты считают наиболее информативными параметрами [Feng: 273]. Nisreen Innab и соавт. (2024) с помощью аналогичного подхода идентифицировали долгосрочную вариабельность как наиболее значимый признак при точности модели 0,9989 [Innab: 1]. Это важно: когда объяснения модели согласуются с клинической логикой, доверие к системе существенно возрастает. Edwin Chandraharan и соавт. (2024) пришли к аналогичному выводу с другой стороны. ИИ-система, обученная на физиологически-ориентированной интерпретации КТГ (то есть на понимании того, какой тип гипоксического стресса стоит за конкретным паттерном), достигала уровня согласия с экспертами ( $\kappa=0,81$ ), заметно превышающего типичное межэкспертное согласие при стандартной категоризации. Это указывает на то, что качество концептуальной основы, заложенной в обучение, не менее важно, чем сложность архитектуры [Chandraharan: 1].

При всём этом разрыв между исследовательским прототипом и реально работающим клиническим инструментом оказывается значительно шире, чем следует из большинства публикаций. Проблема переносимости модели на данные из разных учреждений, популяций и оборудования остаётся нерешённой. Lochana Mendis и соавт. (2025) целенаправленно исследовали этот вопрос, тестируя шесть глубоких нейросетей на двух независимых наборах данных: все модели демонстрировали существенное падение производительности при переносе на внешние данные [Mendis, 2025: 123]. Tae Jun Park и соавт. (2022), обучив модель на 17 492 корейских КТГ с AUC 0,89 на внутренней выборке, получили снижение до 0,73 при тестировании на STU-UHB [Park: 692]. Farah Francis и соавт. (2023) в своих выводах прямо пишут, что для подтверждения обобщаемости необходим внешний датасет. Это редкий пример того, когда авторы сами честно называют этот разрыв [Francis, 2023: 1]. К инфраструктурным ограничениям добавляется и практическое. Значительная часть учреждений по-прежнему использует бумажные КТГ-ленты, тогда как большинство моделей разработаны для цифрового сигнала. Sibel Öztürk и соавт. (2023) специально разработали систему оцифровки бумажных КТГ с последующей ML-классификацией, достигнув точности около 90 %. Это важный шаг к применению в реальной клинической инфраструктуре [Öztürk: 42521]. Наконец, большинство систем анализируют завершённую запись после родов, тогда как клинически значима оценка именно в реальном времени. Lochana Mendis и соавт. (2024) разработали модель, способную прогрессивно оценивать состояние плода по мере развития родов, сократив время обнаружения дистресса примерно на 25 % [Mendis, 2024: 12615].

Единственными системами, преодолевшими этот разрыв в полной мере, остаются коммерческие решения: SisPorto 4.0, обновлённая в соответствии с руководством FIGO-2015

и работающая в режиме реальных предупреждений [Ayres-de-Campos: 62], и qCTG, прошедшая рандомизированное контролируемое испытание с клинически верифицированными исходами (снижение риска гипоксии в 1,9 раза, ацидемии – в 3,2 раза по сравнению с обычным КТГ-мониторингом) [Ignatov: 91]. Схожие результаты на отечественном материале получили М.А. Егорова и соавт. (2017): автоматизированный анализ КТГ по критериям Доуза–Редмана выявлял отклонения у 83 % пациенток группы высокого риска, тогда как визуальная оценка давала значимо более низкие показатели [Егорова: 38]. Эти данные показывают, что переход от исследовательского прототипа к клинически валидированному инструменту принципиально меняет результат. Алгоритм, встроенный в реальный рабочий процесс и проверенный на клинических исходах, работает иначе, чем модель с высокой точностью на конкретном датасете.

Таким образом, алгоритмы машинного обучения доказали способность анализировать КТГ точнее и последовательнее, чем визуальная оценка в ряде хорошо контролируемых условий. Вместе с тем ни одна из исследовательских разработок пока не стала полноценным клиническим инструментом: проблемы несбалансированности данных, недостаточной обобщаемости, малой распространённости использования оцифрованных записей КТГ в медицинских учреждениях и отсутствия режима реального времени остаются нерешёнными для большинства систем. Это определяет направление дальнейших разработок – создание решений, в которых точность алгоритма, его интерпретируемость и встроенность в реальный рабочий процесс клинициста объединены в единую систему.

\* \* \* \* \*

Настоящий систематический обзор выполнен с целью систематизировать данные о методах интерпретации КТГ, выявить ключевые ограничения современных подходов и оценить перспективы применения систем поддержки врачебных решений на основе искусственного интеллекта. Анализ 67 публикаций за период 2015–2025 годов позволил последовательно решить поставленные задачи и сформулировать ряд взаимосвязанных выводов.

Субъективность визуальной интерпретации КТГ подтверждена как устойчивое, воспроизводимое и клинически значимое явление. Ни квалификация специалистов, ни переход к более детализированным классификационным системам не устраняют её в достаточной мере. Коэффициенты межэкспертного согласия при итоговой категоризации КТГ устойчиво остаются на «незначительном» или «справедливом» уровне вне зависимости от используемого руководства. Данные Kundu и соавт. указывают на то, что проблема глубже простого расхождения между специалистами: даже согласованная классификация КТГ слабо коррелирует с прогнозированием реального состояния новорождённого, и доля пропущенных случаев ацидоза достигает 20–70 %. Более того, сравнение пяти действующих классификационных систем показало, что ни одна из них не обеспечивает одновременно приемлемой чувствительности и специфичности: все системы вынуждены выбирать между риском пропустить патологию и риском избыточного вмешательства. Это означает, что совершенствование руководств и обучение специалистов необходимы, но недостаточны.

Гипотеза о том, что автоматизированные методы анализа КТГ способны снизить межэкспертную вариабельность и повысить диагностическую точность, находит частичное подтверждение в литературе. Алгоритмы МО демонстрируют точность от 87 % до 99 % в контролируемых условиях и применяют одни и те же правила последовательно, без влияния усталости или субъективного опыта. Однако большинство результатов получены на одном несбалансированном датасете UCI, и при переносе моделей на данные из других учреждений производительность неизменно снижается. Таким образом, потенциал автоматизации реален, но его реализация в клинической практике требует решения проблемы обобщаемости через

многоцентровую внешнюю валидацию.

Совокупность выявленных ограничений позволяет очертить требования к системам поддержки врачебных решений следующего поколения. Такие системы должны работать с реальной клинической инфраструктурой, включая бумажные КТГ-ленты, выдавать стандартизированное заключение, учитывающее клинический контекст конкретной пациентки, применять одни и те же правила последовательно, устраняя субъективный фактор. Перспективным направлением остаётся многоцентровая внешняя валидация подобных систем на репрезентативных клинических данных и их последующая интеграция в повседневную акушерскую практику.

С учётом полученных данных в рамках настоящего исследования разрабатывается программно-аппаратный комплекс. Предлагаемая схема устройства комплекса предполагает, что врач фотографирует бумажную КТГ-ленту и загружает изображение на веб-сервис; далее снимок поступает на сервер, где модель машинного обучения распознаёт паттерны сигнала и с учётом введённого клинического контекста формирует стандартизированное заключение. Такой подход устраняет необходимость в специализированном цифровом оборудовании и делает инструмент доступным в условиях реальной клинической инфраструктуры, в том числе для пациенток из отдалённых районов, где возможности оперативной консультации профильного специалиста ограничены. Разработка ведётся в рамках Программы поддержки НИОКР студентов, аспирантов и лиц, имеющих учёную степень, финансируемой Правительством Республики Карелия.

### Список литературы

*Егорова М.А.* Возможности визуального и автоматизированного анализа кардиотокограмм в оценке состояния плода при осложненной беременности / М.А. Егорова, А.В. Чурилов, Ю.А. Талалаенко, Е.Д. Минович, Е.В. Литвинова // Медико-социальные проблемы семьи. 2017. Т. 22. № 1. С. 38-43.

*Amadori R.* Intrapartum cardiotocography: an exploratory analysis of interpretational variation / R. Amadori, E. Vaianella, M. Tosi, P. Baronchelli, D. Surico, V. Remorgida // Journal of Obstetrics and Gynaecology. 2022. Vol. 42. № 7. P. 2753-2757. DOI: 10.1080/01443615.2022.2109131

*Asfaw D.* Multimodal deep learning for predicting adverse birth outcomes based on early labour data / D. Asfaw, I. Jordanov, L. Impey, A. Namburete, R. Lee, A. Georgieva // Bioengineering. 2023. Vol. 10. № 6, 730. 17 p. DOI: 10.3390/bioengineering10060730

*Ayres-de-Campos D.* SisPorto 4.0 — computer analysis following the 2015 FIGO guidelines for intrapartum fetal monitoring / D. Ayres-de-Campos, M. Rei, I. Nunes, P. Sousa, J. Bernardes // The Journal of Maternal-Fetal & Neonatal Medicine. 2017. Vol. 30. № 1. P. 62-67. DOI: 10.3109/14767058.2016.1161750

*Ben M'Barek I.* Large-scale analysis of interobserver agreement and reliability in cardiotocography interpretation during labor using an online tool / I. Ben M'Barek, B. Ben M'Barek, G. Jauvion, E. Holmström, A. Agman, J. Merrer, P.-F. Ceccaldi // BMC Pregnancy and Childbirth. 2024. Vol. 24. № 1, 136. DOI: 10.1186/s12884-024-06322-4

*Bhatia M.* A cross-sectional comparison of three guidelines for intrapartum cardiotocography / M. Bhatia, K.R. Mahtani, D. Nunan, A. Reddy // International Journal of Gynaecology and Obstetrics. 2017. Vol. 138. № 1. P. 89-93. DOI: 10.1002/ijgo.12161

*Chandrabaran E.* Physiological interpretation of cardiotocograph (CTG): inter-observer and "observer-AI" agreements in recognising types of fetal hypoxic stress / E. Chandrabaran, K. Griffiths, M. Edmondson, C. Ingram, et al. // Medical Research Archives. 2024. Vol. 12. № 11. DOI: 10.18103/mra.v12i11.5980

*Cömert Z.* A novel software for comprehensive analysis of cardiotocography signals "CTG-OAS" / Z. Cömert, A.F. Kocamaz // 2017 International Artificial Intelligence and Data Processing Symposium (IDAP). 2017. P. 1-6. DOI: 10.1109/IDAP.2017.8090210

*Daydulo Y.D.* Deep learning based fetal distress detection from time frequency representation of cardiotocogram signal using Morse wavelet / Y.D. Daydulo, B.L. Thamineni, H.K. Dasari, G.T. Aboye // BMC Medical Informatics and Decision Making. 2022. Vol. 22. № 1. P. 329. DOI: 10.1186/s12911-022-02068-1

*Feng J.* A hybrid stacked ensemble and kernel SHAP-based model for intelligent cardiotocography classification and interpretability / J. Feng, J. Liang, Z. Qiang, et al. // BMC Medical Informatics and Decision

Making. 2023. Vol. 23. № 1. P. 273. DOI: 10.1186/s12911-023-02378-y

*Francis F.* Detecting intrapartum fetal hypoxia from cardiotocography using machine learning / F. Francis, S. Luz, H. Wu, R. Townsend, S. Stock // 2022 Computing in Cardiology. 2022. Vol. 49. P. 1-4. DOI: 10.22489/CinC.2022.339

*Francis F.* Machine learning to classify cardiotocography for fetal hypoxia detection / F. Francis, S. Luz, H. Wu, R. Townsend, S. Stock // 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2023. P. 1-4. DOI: 10.1109/EMBC40787.2023.10340803

*Garabedian C.* Inter-observer reliability of 4 fetal heart rate classifications / C. Garabedian, L. Butruille, E. Drumez, et al. // Journal of Gynecology Obstetrics and Human Reproduction. 2017. Vol. 46. № 2. P. 131-135. DOI: 10.1016/j.jogoh.2016.11.002

*Hruban L.* Agreement on intrapartum cardiotocogram recordings between expert obstetricians / L. Hruban, J. Spilka, V. Chudáček, et al. // Journal of Evaluation in Clinical Practice. 2015. Vol. 21. № 4. P. 694-702. DOI: 10.1111/jep.12368

*Ignatov P.N.* Quantitative cardiotocography to improve fetal assessment during labor: a preliminary randomized controlled trial / P.N. Ignatov, J.E. Lutomski // European Journal of Obstetrics, Gynecology, and Reproductive Biology. 2016. Vol. 205. P. 91-97. DOI: 10.1016/j.ejogrb.2016.08.023

*Innab N.* Automated approach for fetal and maternal health management using light gradient boosting model with SHAP explainable AI / N. Innab, S. Alsubai, E.A. Alabdulqader, et al. // Frontiers in Public Health. 2024. Vol. 12. P. 1462693. DOI: 10.3389/fpubh.2024.1462693

*Khan M.J.* PatchCTG: a patch cardiotocography transformer for antepartum fetal health monitoring / M.J. Khan, M. Vatish, G. Davis Jones // Sensors. 2025. Vol. 25. № 9. P. 2650. DOI: 10.3390/s25092650

*Kling D.* Performance of five cardiotocography classification templates in labor: a cohort study / D. Kling, M. Rehnström, A. Herbst // The Journal of Maternal-Fetal & Neonatal Medicine. 2024. Vol. 37. № 1. P. 2394845. DOI: 10.1080/14767058.2024.2394845

*Kundu S.* Estimation of neonatal outcome artery pH value according to CTG interpretation of the last 60 min before delivery: a retrospective study / S. Kundu, E. Kuehnle, C. Schippert, et al. // Archives of Gynecology and Obstetrics. 2017. Vol. 296. № 5. P. 897-905. DOI: 10.1007/s00404-017-4516-4

*Ma'sum M.A.* Improving deep learning classifier for fetus hypoxia detection in cardiotocography signal / M.A. Ma'sum, P. Riskyana Dewi Intan, W. Jatmiko, et al. // 2019 International Workshop on Big Data and Information Security (IWBIS). 2019. P. 51-56. DOI: 10.1109/IWBIS.2019.8935835

*Martí Gamboa S.* Diagnostic accuracy of the FIGO and the 5-tier fetal heart rate classification systems in the detection of neonatal acidemia / S. Martí Gamboa, O.R. Giménez, J.P. Mancho, et al. // American Journal of Perinatology. 2017. Vol. 34. № 5. P. 508-514. DOI: 10.1055/s-0036-1593810

*Mendis L.* Rapid detection of fetal compromise using input length invariant deep learning on fetal heart rate signals / L. Mendis, M. Palaniswami, E. Keenan, F. Brownfoot // Scientific Reports. 2024. Vol. 14. № 1. P. 12615. DOI: 10.1038/s41598-024-63108-6

*Mendis L.* Cross-database evaluation of deep learning methods for intrapartum cardiotocography classification / L. Mendis, D. Karmakar, M. Palaniswami, F. Brownfoot, E. Keenan // IEEE Journal of Translational Engineering in Health and Medicine. 2025. Vol. 13. P. 123-135. DOI: 10.1109/JTEHM.2025.3548401

*Nagendra V.* Evaluation of support vector machines and random forest classifiers in a real-time fetal monitoring system based on cardiotocography data / V. Nagendra, H. Gude, D. Sampath, S. Corns, S. Long // 2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). 2017. P. 1-6. DOI: 10.1109/CIBCB.2017.8058546

*Nazli I.* Early detection of fetal health conditions using machine learning for classifying imbalanced cardiotocographic data / I. Nazli, E. Korbeko, S. Dogru, E. Kugu, O.K. Sahingoz // Diagnostics. 2025. Vol. 15. № 10. P. 1250. DOI: 10.3390/diagnostics15101250

*Neri S.* Interobserver agreement of intrapartum cardiotocography interpretation by midwives using current FIGO and physiology-based guidelines / S. Neri, R. Ramirez Zegarra, M. Dininno, et al. // The Journal of Maternal-Fetal & Neonatal Medicine. 2024. Vol. 37. № 1. P. 2425758. DOI: 10.1080/14767058.2024.2425758

*Neri S.* Interobserver agreement among midwives in cardiotocography interpretation using the 2015 FIGO guidelines for intrapartum fetal monitoring: a retrospective study / S. Neri, R. Ramirez Zegarra, E. Nicolini, et al. // International Journal of Gynecology & Obstetrics. 2025. P. 70687. DOI: 10.1002/ijgo.70687

*Ogasawara J.* Deep neural network-based classification of cardiotocograms outperformed conventional algorithms / J. Ogasawara, S. Ikenoue, H. Yamamoto, et al. // Scientific Reports. 2021. Vol. 11. № 1. P. 13367.

DOI: 10.1038/s41598-021-92805-9

Öztürk S. A novel approach for cardiotocography paper digitization and classification for abnormality detection / S. Öztürk, S.A. Şahin, A.N. Aksoy, et al. // IEEE Access. 2023. Vol. 11. P. 42521-42533. DOI: 10.1109/ACCESS.2023.3271137

Park T.J. Machine learning model for classifying the results of fetal cardiotocography conducted in high-risk pregnancies / T.J. Park, H.J. Chang, B.J. Choi, et al. // Yonsei Medical Journal. 2022. Vol. 63. № 7. P. 692-700. DOI: 10.3349/ymj.2022.63.7.692

Petrozziello A. Deep learning for continuous electronic fetal monitoring in labor / A. Petrozziello, I. Jordanov, A.T. Papageorghiou, C.W.G. Redman, A. Georgieva // Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2018. P. 5866-5869. DOI: 10.1109/EMBC.2018.8513625

Petrozziello A. Multimodal convolutional neural networks to detect fetal compromise during labor and delivery / A. Petrozziello, C.W.G. Redman, A.T. Papageorghiou, I. Jordanov, A. Georgieva // IEEE Access. 2019. Vol. 7. P. 112026-112036. DOI: 10.1109/ACCESS.2019.2933368

Pradhan A.K. A machine learning approach for the prediction of fetal health using CTG / A.K. Pradhan, J.K. Rout, A.B. Maharana, et al. // 2021 19th OITS International Conference on Information Technology (OCIT). 2021. P. 239-244. DOI: 10.1109/OCIT53463.2021.00056

Rei M. Interobserver agreement in CTG interpretation using the 2015 FIGO guidelines for intrapartum fetal monitoring / M. Rei, S. Tavares, P. Pinto, et al. // European Journal of Obstetrics & Gynecology and Reproductive Biology. 2016. Vol. 205. P. 27-31. DOI: 10.1016/j.ejogrb.2016.08.017

Rovas L. Interobserver agreement: individual CTG features show better agreement among investigators than the overall CTG assessment in cases of meconium-stained amniotic fluid / L. Rovas, M. Minkauskiene, K. Berskiene, et al. // European Journal of Midwifery. 2025. Vol. 9. P. 55. DOI: 10.18332/ejm/215682

Sabiani L. Intra- and interobserver agreement among obstetric experts in court regarding the review of abnormal fetal heart rate tracings and obstetrical management / L. Sabiani, R. Le Dû, A. Loundou, et al. // American Journal of Obstetrics and Gynecology. 2015. Vol. 213. № 6. P. 856.e1-856.e8. DOI: 10.1016/j.ajog.2015.08.066

Shastry K.A. Adaptive weighted diversity ensemble learning approach for fetal health classification on cardiotocography data / K.A. Shastry, M.S. Gounder, T.G.M. Kumar, et al. // IEEE Access. 2024. Vol. 12. P. 190690-190710. DOI: 10.1109/ACCESS.2024.3516880

Vejux N. Guideline choice for CTG analysis influences first caesarean decision / N. Vejux, R. Ledu, C. D'Ercole, et al. // The Journal of Maternal-Fetal & Neonatal Medicine. 2017. Vol. 30. № 15. P. 1816-1819. DOI: 10.1080/14767058.2016.1228050

Wong W.K. Fetal health prediction from cardiotocography recordings using Kolmogorov–Arnold networks / W.K. Wong, F.H. Juwono, C. Apriono, I.R. Fitri // IEEE Open Journal of Engineering in Medicine and Biology. 2025. Vol. 6. P. 345-351. DOI: 10.1109/OJEMB.2025.3549594

Zamora del Pozo C. Diagnostic capacity and interobserver variability in FIGO, ACOG, NICE and Chandrachan cardiotocographic guidelines to predict neonatal acidemia / C. Zamora del Pozo, M. Chóliz Ezquerro, I. Mejía, et al. // The Journal of Maternal-Fetal & Neonatal Medicine. 2022. Vol. 35. № 25. P. 8498-8506. DOI: 10.1080/14767058.2021.1986479

Ulyana E. TAEVA

specialist degree, Petrozavodsk State University  
(Petrozavodsk, Russia),  
[taevauliana@gmail.com](mailto:taevauliana@gmail.com)

## SUBJECTIVITY IN VISUAL CARDIOTOCOGRAPHY INTERPRETATION AND PROSPECTS FOR MACHINE-BASED ANALYSIS

**Scientific adviser:**

Alexander A. Ivshin

**Reviewer:**

Julia Boldina

Paper submitted on: 05/07/2026;

Accepted on: 06/27/2026;

Published online on: 06/27/2026.

**Abstract.** Cardiotocography is the primary method of intrauterine foetal monitoring. However, its visual interpretation is subjective: inter-expert agreement on assessment does not exceed 48 %, and no guidelines provide both high sensitivity and specificity simultaneously. A review of 67 publications between 2015 and 2025 showed that ML algorithms outperform visual assessment in terms of reproducibility, achieving accuracy of 87–99 %, but their applicability is limited by training data imbalance and reduced accuracy on new samples. A promising direction is integrating AI into clinical decision support systems.

**Keywords:** cardiotocography, machine learning, inter-expert variability, clinical decision support system, CTG interpretation, foetal monitoring, artificial intelligence

**For citation:** Taeva, U. E. Subjectivity in Visual Cardiotocography Interpretation and Prospects for Machine-Based Analysis. *StudArctic Forum*. 2026, 11 (2): 42–55.

### References

- Egorova M.A., Churilov A.V., et al. Possibilities of visual and automated cardiotocogram analysis for assessment of fetal state in complicated pregnancy. *Mediko-sotsialniye problemy semyi*, 2017, Vol. 22, No. 1, pp. 38–43. (In Russ.)
- Amadori R., Vaianella E., et al. Intrapartum cardiotocography: an exploratory analysis of interpretational variation. *Journal of Obstetrics and Gynaecology*, 2022, Vol. 42, № 7, pp. 2753–2757. DOI: 10.1080/01443615.2022.2109131
- Asfaw D., Jordanov I., et al. Multimodal deep learning for predicting adverse birth outcomes based on early labour data. *Bioengineering*, 2023, Vol. 10, № 6, 730. DOI: 10.3390/bioengineering10060730
- Ayres-de-Campos D., Rei M., et al. SisPorto 4.0 — computer analysis following the 2015 FIGO guidelines for intrapartum fetal monitoring. *The Journal of Maternal-Fetal & Neonatal Medicine*, 2017, Vol. 30, No. 1, pp. 62–67. DOI: 10.3109/14767058.2016.1161750
- Ben M'Barek I., Ben M'Barek B., et al. Large-scale analysis of interobserver agreement and reliability in cardiotocography interpretation during labor using an online tool. *BMC Pregnancy and Childbirth*, 2024, Vol. 24, No. 1, pp. 136. DOI: 10.1186/s12884-024-06322-4
- Bhatia M., Mahtani K. R., et al. A cross-sectional comparison of three guidelines for intrapartum cardiotocography. *International Journal of Gynaecology and Obstetrics*, 2017, Vol. 138, No. 1, pp. 89–93. DOI: 10.1002/ijgo.12161
- Chandrabaran E., Griffiths K., et al. Physiological interpretation of cardiotocograph (CTG): inter-observer and "observer-AI" agreements in recognising types of fetal hypoxic stress. *Medical Research Archives*, 2024, Vol. 12, No. 11. DOI: 10.18103/mra.v12i11.5980
- Cömert Z., Kocamaz A. F. A novel software for comprehensive analysis of cardiotocography signals "CTG-OAS". *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*. 2017, pp. 1–6. DOI: 10.1109/IDAP.2017.8090210
- Daydulo Y.D., Thamineni B.L., et al. Deep learning based fetal distress detection from time frequency representation of cardiotocogram signal using Morse wavelet. *BMC Medical Informatics and Decision Making*, 2022, Vol. 22, No. 1, pp. 329. DOI: 10.1186/s12911-022-02068-1
- Feng J., Liang J., et al. A hybrid stacked ensemble and kernel SHAP-based model for intelligent

cardiotocography classification and interpretability. *BMC Medical Informatics and Decision Making*, 2023, Vol. 23, No. 1, p. 273. DOI: 10.1186/s12911-023-02378-y

Francis F., Luz S., Wu H., et al. Detecting intrapartum fetal hypoxia from cardiotocography using machine learning. *2022 Computing in Cardiology*, 2022, Vol. 49, pp. 1–4. DOI: 10.22489/CinC.2022.339

Francis F., Luz S., et al. Machine learning to classify cardiotocography for fetal hypoxia detection. *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 2023, pp. 1–4. DOI: 10.1109/EMBC40787.2023.10340803

Garabedian C., Butruille L., et al. Inter-observer reliability of 4 fetal heart rate classifications. *Journal of Gynecology Obstetrics and Human Reproduction*, 2017, Vol. 46, No. 2, pp. 131–135. DOI: 10.1016/j.jogoh.2016.11.002

Hruban L., Spilka J., et al. Agreement on intrapartum cardiotocogram recordings between expert obstetricians. *Journal of Evaluation in Clinical Practice*, 2015, Vol. 21, No. 4, pp. 694–702. DOI: 10.1111/jep.12368

Ignatov P.N., Lutomski J.E. Quantitative cardiotocography to improve fetal assessment during labor: a preliminary randomized controlled trial. *European Journal of Obstetrics, Gynecology, and Reproductive Biology*, 2016, Vol. 205, pp. 91–97. DOI: 10.1016/j.ejogrb.2016.08.023

Innab N., Alsubai S., et al. Automated approach for fetal and maternal health management using light gradient boosting model with SHAP explainable AI. *Frontiers in Public Health*, 2024, Vol. 12, pp. 1462693. DOI: 10.3389/fpubh.2024.1462693

Khan M.J., Vatish M., et al. PatchCTG: a patch cardiotocography transformer for antepartum fetal health monitoring. *Sensors*, 2025, Vol. 25, No. 9, pp. 2650. DOI: 10.3390/s25092650

Kling D., Rehnström M., et al. Performance of five cardiotocography classification templates in labor: a cohort study. *The Journal of Maternal-Fetal & Neonatal Medicine*, 2024, Vol. 37, No. 1, pp. 2394845. DOI: 10.1080/14767058.2024.2394845

Kundu S., Kuehne E., et al. Estimation of neonatal outcome artery pH value according to CTG interpretation of the last 60 min before delivery: a retrospective study. *Archives of Gynecology and Obstetrics*, 2017, Vol. 296, No. 5, pp. 897–905. DOI: 10.1007/s00404-017-4516-4

Ma'sum M.A., Riskyana Dewi Intan P., et al. Improving deep learning classifier for fetus hypoxia detection in cardiotocography signal. *2019 International Workshop on Big Data and Information Security (IW BIS)*. 2019, pp. 51–56. DOI: 10.1109/IWBIS.2019.8935835

Martí Gamboa S., Giménez O.R., et al. Diagnostic accuracy of the FIGO and the 5-tier fetal heart rate classification systems in the detection of neonatal acidemia. *American Journal of Perinatology*, 2017, Vol. 34, No. 5, pp. 508–514. DOI: 10.1055/s-0036-1593810

Mendis L., Palaniswami M., et al. Rapid detection of fetal compromise using input length invariant deep learning on fetal heart rate signals. *Scientific Reports*, 2024, Vol. 14, No. 1, pp. 12615. DOI: 10.1038/s41598-024-63108-6

Mendis L., Karmakar D., et al. Cross-database evaluation of deep learning methods for intrapartum cardiotocography classification. *IEEE Journal of Translational Engineering in Health and Medicine*, 2025, Vol. 13, pp. 123–135. DOI: 10.1109/JTEHM.2025.3548401

Nagendra V., Gude H., et al. Evaluation of support vector machines and random forest classifiers in a real-time fetal monitoring system based on cardiotocography data. *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. 2017, pp. 1–6. DOI: 10.1109/CIBCB.2017.8058546

Nazli I., Korbeko E., et al. Early detection of fetal health conditions using machine learning for classifying imbalanced cardiotocographic data. *Diagnostics*, 2025, Vol. 15, No. 10, pp. 1250. DOI: 10.3390/diagnostics15101250

Neri S., Ramirez Zegarra R., et al. Interobserver agreement of intrapartum cardiotocography interpretation by midwives using current FIGO and physiology-based guidelines. *The Journal of Maternal-Fetal & Neonatal Medicine*, 2024, Vol. 37, No. 1, pp. 2425758. DOI: 10.1080/14767058.2024.2425758

Neri S., Ramirez Zegarra R., et al. Interobserver agreement among midwives in cardiotocography interpretation using the 2015 FIGO guidelines for intrapartum fetal monitoring: a retrospective study. *International Journal of Gynecology & Obstetrics*, 2025, p. 70687. DOI: 10.1002/ijgo.70687

Ogasawara J., Ikenoue S., et al. Deep neural network-based classification of cardiotocograms outperformed conventional algorithms. *Scientific Reports*, 2021, Vol. 11, No. 1, p. 13367. DOI: 10.1038/s41598-021-92805-9

- Öztürk S., Şahin S.A., et al. A novel approach for cardiotocography paper digitization and classification for abnormality detection. *IEEE Access*, 2023, Vol. 11, pp. 42521–42533. DOI: 10.1109/ACCESS.2023.3271137
- Park T.J., Chang H.J., et al. Machine learning model for classifying the results of fetal cardiotocography conducted in high-risk pregnancies. *Yonsei Medical Journal*, 2022, Vol. 63, No. 7, pp. 692–700. DOI: 10.3349/ymj.2022.63.7.692
- Petrozziello A., Jordanov I., et al. Deep learning for continuous electronic fetal monitoring in labor. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2018, pp. 5866–5869. DOI: 10.1109/EMBC.2018.8513625
- Petrozziello A., Redman C.W.G., et al. Multimodal convolutional neural networks to detect fetal compromise during labor and delivery. *IEEE Access*, 2019, Vol. 7, pp. 112026–112036. DOI: 10.1109/ACCESS.2019.2933368
- Pradhan A.K., Rout J.K., et al. A machine learning approach for the prediction of fetal health using CTG. *2021 19th OITS International Conference on Information Technology (OCIT)*. 2021, pp. 239–244. DOI: 10.1109/OCIT53463.2021.00056
- Rei M., Tavares S., et al. Interobserver agreement in CTG interpretation using the 2015 FIGO guidelines for intrapartum fetal monitoring. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 2016, Vol. 205, pp. 27–31. DOI: 10.1016/j.ejogrb.2016.08.017
- Rovas L., Minkauskiene M., et al. Interobserver agreement: individual CTG features show better agreement among investigators than the overall CTG assessment in cases of meconium-stained amniotic fluid. *European Journal of Midwifery*, 2025, Vol. 9, pp. 55. DOI: 10.18332/ejm/215682
- Sabiani L., Le Dû R., et al. Intra- and interobserver agreement among obstetric experts in court regarding the review of abnormal fetal heart rate tracings and obstetrical management. *American Journal of Obstetrics and Gynecology*, 2015, Vol. 213, No. 6, pp. 856.e1–856.e8. DOI: 10.1016/j.ajog.2015.08.066
- Shastry K.A., Gounder M.S., et al. Adaptive weighted diversity ensemble learning approach for fetal health classification on cardiotocography data. *IEEE Access*, 2024, Vol. 12, pp. 190690–190710. DOI: 10.1109/ACCESS.2024.3516880
- Vejux N., Ledu R., et al. Guideline choice for CTG analysis influences first caesarean decision. *The Journal of Maternal-Fetal & Neonatal Medicine*, 2017, Vol. 30, No. 15, pp. 1816–1819. DOI: 10.1080/14767058.2016.1228050
- Wong W.K., Juwono F.H., et al. Fetal health prediction from cardiotocography recordings using Kolmogorov–Arnold networks. *IEEE Open Journal of Engineering in Medicine and Biology*, 2025, Vol. 6, pp. 345–351. DOI: 10.1109/OJEMB.2025.3549594
- Zamora del Pozo C., Chóliz Ezquerro M., et al. Diagnostic capacity and interobserver variability in FIGO, ACOG, NICE and Chandrachud cardiotocographic guidelines to predict neonatal acidemia. *The Journal of Maternal-Fetal & Neonatal Medicine*, 2022, Vol. 35, No. 25, pp. 8498–8506. DOI: 10.1080/14767058.2021.1986479