doi.org/10.51891/rease.v12i11.31754

PAPEL DA INTELIGÊNCIA ARTIFICIAL NA DETECÇÃO AUTOMÁTICA DE MALFORMAÇÕES FETAIS POR ULTRASSONOGRAFIA

ROLE OF ARTIFICIAL INTELLIGENCE IN THE AUTOMATIC DETECTION OF FETAL MALFORMATIONS BY ULTRASOUND

PAPEL DE LA INTELIGENCIA ARTIFICIAL EN LA DETECCIÓN AUTOMÁTICA DE MALFORMACIONES FETALES POR ULTRASONOGRAFÍA

Eduarda Fonseca Martins[1]
Fernanda Medeiros Andrade Linhares[2]
Jordana de Oliveira Monteiro[3]
Maria Eduarda Viana Cirilo[4]
Márcio José Rosa Requeijo[5]

RESUMO: A ultrassonografia morfológica permanece o principal instrumento de rastreamento de malformações fetais, com sensibilidade que varia amplamente conforme a experiência do examinador e a qualidade da aquisição. Esta revisão narrativa crítica examina o papel dos sistemas de inteligência artificial na detecção automática de anomalias fetais por ultrassonografia, separando o que já foi demonstrado em validação clínica daquilo que permanece como promessa técnica. O levantamento foi realizado em 2026 na PubMed/MEDLINE, complementado por rastreamento das listas de referências e por consulta a documentos regulatórios oficiais. A análise percorre a cadeia diagnóstica: aquisição e controle de qualidade da imagem, desempenho por território anatômico, efeito da assistência sobre o examinador, generalização e condições de incorporação. Alguns modelos alcançam, em tarefas circunscritas e sobretudo no rastreamento cardíaco e neurossonográfico, acurácia próxima à de especialistas, e elevam o desempenho de operadores menos experientes; a heterogeneidade entre estudos, os comparadores humanos reduzidos e a validação externa insuficiente impedem generalizar esse desempenho. Persistem a dependência da qualidade da imagem, a escassez de validação externa prospectiva e o risco de deferência a sugestões incorretas. O ganho plausível é de padronização e de resgate de casos perdidos por falha de aquisição, não de substituição do examinador.

Palavras-chave: Inteligência Artificial. Ultrassonografia Pré-Natal. Anormalidades Congênitas.

[1]Faminas-BH.
[2]Faminas-BH.
[3]Faminas-BH.
[4]Faminas-BH.
[5]Orientador. Faminas-BH.

ABSTRACT: Anatomical ultrasound remains the main screening tool for fetal malformations, although its sensitivity varies widely according to operator experience and image acquisition quality. This critical narrative review examines the role of artificial intelligence systems in the automatic detection of fetal anomalies by ultrasound, aiming to separate what has already been demonstrated in clinical validation from what remains a technical promise. The literature search was conducted in 2026, with PubMed/MEDLINE as the primary database, complemented by reference list screening and direct consultation of official regulatory documents. The analysis follows the diagnostic chain: image acquisition and quality control, performance by anatomical territory, the effect of assistance on the examiner, generalization, and conditions for adoption. Some models achieve accuracy close to that of specialists in circumscribed tasks, particularly in cardiac and neurosonographic screening, and improve the performance of less experienced operators; heterogeneity across studies, small human comparator groups, and insufficient external validation preclude generalizing this performance. Dependence on image quality, scarcity of prospective external validation, and risk of deference to incorrect suggestions persist. The plausible gain lies in standardization and in the recovery of cases missed through acquisition failure, not in replacing the examiner.

Keywords: Artificial Intelligence. Ultrasonography. Prenatal. Congenital Abnormalities.

RESUMEN: La ultrasonografía morfológica sigue siendo el principal instrumento de cribado de malformaciones fetales, con una sensibilidad que varía ampliamente según la experiencia del examinador y la calidad de la adquisición. Esta revisión narrativa crítica examina el papel de los sistemas de inteligencia artificial en la detección automática de anomalías fetales por ultrasonografía, separando lo que ya se ha demostrado en validación clínica de lo que permanece como promesa técnica. La búsqueda se realizó en 2026 en PubMed/MEDLINE, complementada con el rastreo de las listas de referencias y la consulta de documentos regulatorios oficiales. El análisis recorre la cadena diagnóstica: adquisición y control de calidad de la imagen, desempeño por territorio anatómico, efecto de la asistencia sobre el examinador, generalización y condiciones de incorporación. Algunos modelos alcanzan, en tareas circunscritas y sobre todo en el cribado cardíaco y neurosonográfico, una exactitud cercana a la de los especialistas, y elevan el desempeño de los operadores menos experimentados; la heterogeneidad entre estudios, los comparadores humanos reducidos y la validación externa insuficiente impiden generalizar ese desempeño. Persisten la dependencia de la calidad de la imagen, la escasez de validación externa prospectiva y el riesgo de deferencia ante sugerencias incorrectas. La ganancia plausible es de estandarización y de rescate de casos perdidos por fallas de adquisición, no de sustitución del examinador.

Palabras clave: Inteligencia Artificial. Ultrasonografía Prenatal. Anomalías Congénitas.

INTRODUÇÃO

Os cinco planos cardíacos previstos no rastreamento ultrassonográfico permitem, em tese, identificar cerca de 90% das cardiopatias congênitas complexas; na prática assistencial, a sensibilidade descrita chega a ser de apenas 30% (ARNAOUT R, et al., 2021). A distância entre o que o protocolo promete e o que o serviço entrega não é uma peculiaridade da cardiologia fetal.

Metanálise de estudos populacionais estimou detecção pré-natal agregada de 45,1% (IC 95%: 33,5%-57,0%) para cardiopatias congênitas em populações não selecionadas, com taxas superiores a 85% apenas para defeitos univentriculares e heterotaxias, e desempenho fortemente correlacionado à gravidade da lesão (VAN VELZEN CL, et al., 2018). No Brasil, coorte conduzida em centro de referência do Nordeste encontrou sensibilidade de 61,1% e especificidade de 96,3% do exame morfológico de segundo trimestre para malformações em geral (LEIROZ R, et al., 2021), enquanto série do Sul do país registrou sensibilidade de 29,3% da ultrassonografia obstétrica para cardiopatias, contra 97,7% da ecocardiografia fetal dirigida (PINHEIRO DO, et al., 2019).

O conteúdo mínimo do exame de segundo trimestre está normatizado há anos, com definição explícita de planos, estruturas e imagens a documentar (SALOMON LJ, et al., 2022). O que varia é a execução. Análise automatizada de 496 exames morfológicos gravados em vídeo mostrou que nenhum deles seguiu sequência idêntica de aquisição, e que as cinco regiões anatômicas de registro obrigatório no protocolo britânico foram documentadas em proporções desiguais, de 86,1% para o plano coronal da face a 97,2% para o polo cefálico (DRUKKER L, et al., 2022b). A variabilidade do gesto técnico precede qualquer erro de interpretação, e é ela que fixa o teto do rastreamento.

Sistemas de aprendizado profundo reconhecem padrões em imagens com desempenho que, em tarefas definidas, aproxima-se do humano, o que explica o crescimento acelerado da produção sobre ultrassonografia obstétrica assistida (RAMIREZ ZEGARRA R e GHI T, 2023). Duas revisões de escopo sucessivas mapearam esse campo e mostraram concentração temática: biometria fetal responde por cerca de um quarto das publicações e a avaliação anatômica do feto por um quinto, com predomínio de estudos conduzidos na Ásia e na Europa e apenas 22% originados em países de renda baixa ou média (HORGAN R, et al., 2023; GUPTA V, et al., 2026). A extração quantitativa de descritores de textura e forma, herdada da radiômica, amplia ainda mais o repertório de tarefas possíveis (BOUACHBA A, et al., 2025).

O entusiasmo, porém, corre à frente da evidência clínica. Métricas obtidas em conjuntos curados, com prevalência artificialmente elevada e comparadores humanos escassos, não se convertem automaticamente em ganho assistencial, e o arcabouço regulatório que disciplinará esses produtos já está posto, tanto no Brasil (BRASIL, 2022) quanto na União Europeia (UNIÃO EUROPEIA, 2024). Este artigo pergunta em que condições a assistência algorítmica

modifica efetivamente a detecção de malformações fetais no rastreamento ultrassonográfico e o que ainda separa o desempenho relatado do benefício clínico demonstrável. O objetivo é examinar criticamente a evidência disponível ao longo da cadeia diagnóstica, distinguindo desempenho técnico em conjuntos de dados de desempenho clínico em condições reais de triagem, e explicitar os requisitos de validação, equidade e regulação necessários a uma incorporação segura.

METODOLOGIA

Trata-se de revisão narrativa crítica da literatura, delineamento escolhido pela natureza heterogênea do material disponível, que reúne estudos de acurácia diagnóstica, ensaios de leitura assistida, análises econômicas e normas regulatórias, sem denominador metodológico comum que autorizasse agregação quantitativa.

A pergunta orientadora foi assim formulada: em que condições a assistência por inteligência artificial altera a detecção de malformações fetais no rastreamento ultrassonográfico, e quais lacunas ainda separam o desempenho relatado do benefício clínico demonstrável?

As buscas foram realizadas em 2026. A base principal foi a PubMed/MEDLINE, escolhida por indexar tanto a literatura de medicina fetal quanto os periódicos de informática biomédica em que os modelos são originalmente descritos, e por dar acesso aos periódicos brasileiros pertinentes ao recorte. A ela somaram-se o rastreamento das listas de referências dos artigos recuperados e a consulta direta aos portais oficiais da Agência Nacional de Vigilância Sanitária e do EUR-Lex, para obtenção do texto integral das normas citadas. Uma formulação representativa da estratégia principal, com termos restritos a título e resumo, foi: (artificial intelligence OR deep learning OR machine learning OR convolutional neural network) AND (fetal OR prenatal) AND (ultrasound OR ultrasonography) AND (anomaly OR malformation OR defect OR screening). Buscas dirigidas complementaram a estratégia para territórios anatômicos específicos, para ensaios de interação entre operador e algoritmo e para diretrizes de relato.

Não foi imposto limite inferior de ano de publicação. A leitura privilegiou o período em que redes neurais convolucionais aplicadas à ultrassonografia obstétrica passaram a ser avaliadas contra desfechos clínicos, de 2020 em diante, admitindo material anterior quando

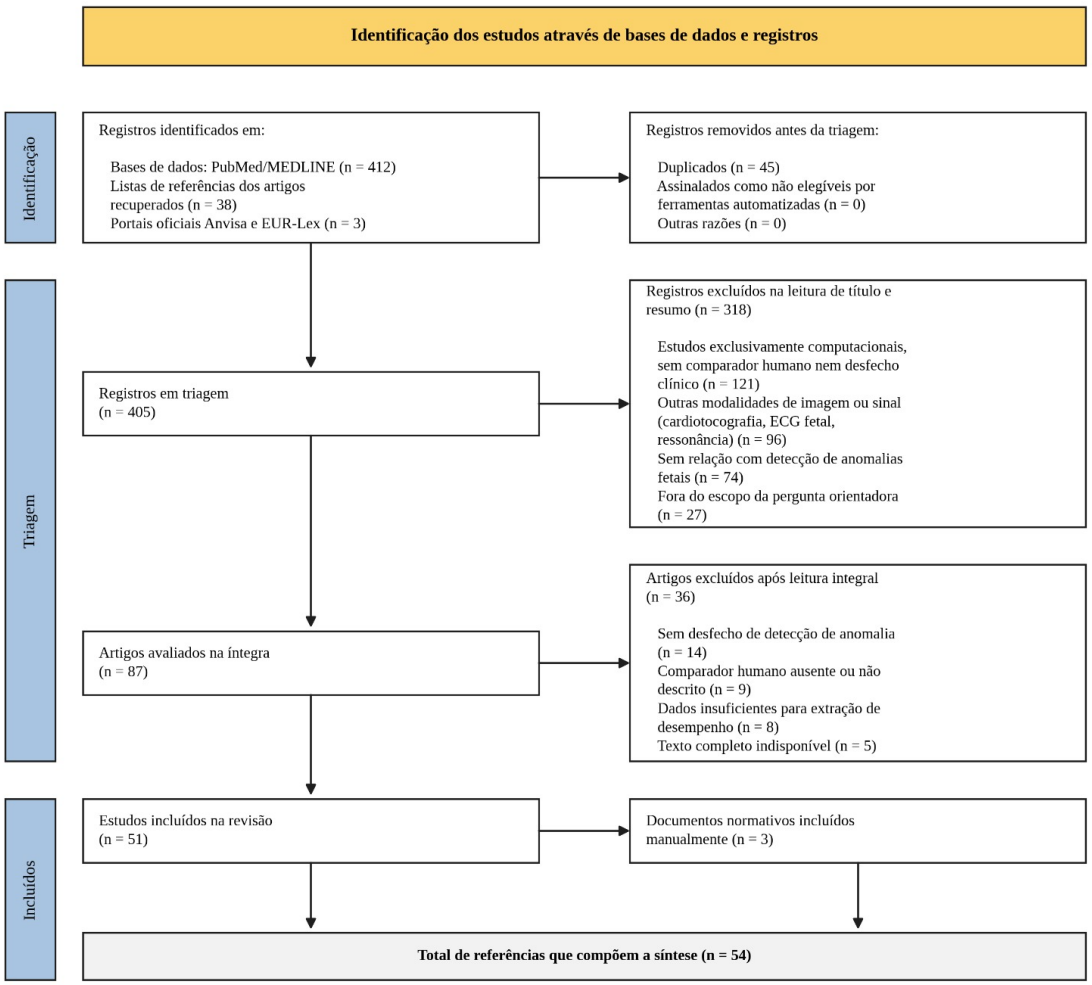
estabelecia a linha de base populacional da detecção pré-natal ou fixava padrão de relato ainda vigente.

Foram considerados elegíveis estudos que avaliassem modelos aplicados a imagens ou vídeos de ultrassonografia obstétrica com desfecho de detecção de anomalia, ou tarefas de reconhecimento de planos e controle de qualidade diretamente subordinadas a essa finalidade; estudos populacionais que informassem taxas de detecção sem assistência algorítmica; análises de custo-utilidade; e diretrizes de relato e normas sanitárias aplicáveis a software como dispositivo médico. Ficaram de fora trabalhos exclusivamente computacionais, sem comparador humano nem desfecho clínico, e aplicações a outras modalidades, como cardiotocografia, eletrocardiografia fetal e ressonância magnética, cujas propriedades de sinal não são transferíveis para a discussão pretendida. Estudos com resultados nulos ou desfavoráveis foram deliberadamente mantidos (Figura 1).

A organização da síntese seguiu a cadeia diagnóstica, e não a ordem cronológica ou a lista de estudos: aquisição da imagem, classificação por território anatômico, interação entre examinador e algoritmo, generalização entre populações e incorporação. Cada métrica relatada foi lida contra o desenho que a produziu, atentando à prevalência da condição no conjunto de teste, à existência de validação externa, à natureza do comparador humano e ao caráter retrospectivo ou prospectivo da avaliação. A seleção não foi exaustiva e não houve avaliação formal de risco de viés, o que impõe leitura cautelosa das comparações entre estudos.

Parte do material discutido é indireta em relação à pergunta orientadora. Estudos de estimativa automatizada da idade gestacional e evidências provenientes de outros domínios de imagem foram empregados apenas para examinar transferência de domínio, infraestrutura de aquisição, generalização e equidade, e não como demonstração de eficácia na detecção de malformações fetais; a distinção foi adotada no curso da análise, e não pré-especificada, e está explicitada nas passagens em que esse material aparece. A opção por uma base biomédica principal, ainda que complementada pelo rastreamento das listas de referências, pode ter reduzido a cobertura da literatura publicada em veículos de engenharia e de ciência da computação, em que parte do desenvolvimento metodológico é divulgada.

Figura 1 - Fluxograma de identificação, triagem e inclusão dos estudos



Fonte: Autoria própria.

AQUISIÇÃO DA IMAGEM: ONDE O RASTREAMENTO COMEÇA A FALHAR

Atribuir a perda diagnóstica à leitura do examinador desloca o problema. Quando 66 casos de cardiopatia isolada grave nascidos em uma região dos Países Baixos foram reanalisados, 31 haviam passado despercebidos no exame morfológico de rotina, o que corresponde a sensibilidade clínica de 53%. Aplicado às mesmas imagens armazenadas, um modelo previamente treinado em centro terciário alcançou sensibilidade de 91% e especificidade de 78%, enquanto três especialistas cegados, com acesso apenas ao mesmo material estático, obtiveram sensibilidade média de 55% (ATHALYE C, et al., 2024). Dois achados desse estudo merecem atenção. O acerto do modelo variou significativamente conforme a qualidade da imagem julgada por avaliadores independentes, e o desempenho em 19 lesões nunca vistas durante o

treinamento não diferiu estatisticamente do observado em lesões conhecidas, o que sugere aprendizado de desvios do padrão anatômico normal, e não apenas memorização de categorias diagnósticas.

Se o insumo limita o resultado, faz sentido intervir na captura. Sistemas de orientação em tempo real avaliam a qualidade do plano enquanto a sonda ainda está sobre o abdome materno e sinalizam estruturas ausentes. Um desses sistemas, treinado com 134.696 imagens e validado em 836 vídeos de quatro hospitais, reconheceu 19 planos padronizados com área sob a curva macro de 0,958 na validação externa e rotulou corretamente 90,9% dos planos quase padronizados, indicando o que faltava para completá-los (LEI T, et al., 2024a). A lógica é de controle de processo, não de diagnóstico: reduzir a proporção de exames que chegam à etapa interpretativa com material insuficiente.

Evidência de mundo real sustenta a plausibilidade dessa via. Avaliação multicêntrica conduzida em 34 hospitais da província de Guizhou, na China, em ambiente de poucos recursos, acompanhou 61.959 exames e 551.144 imagens enviadas por 186 ecografistas a um sistema independente de controle de qualidade por inteligência artificial, que pontua cada imagem e a classifica como padronizada, básica ou não padronizada. Ao longo de 36 meses, a proporção de exames integralmente padronizados aumentou em todas as categorias, do rastreamento de primeiro trimestre, de 39,5% para 82,1%, ao exame morfológico completo de 23 planos, de 16,1% para 53,3%, com ganho sustentado no tempo (ZHAO J, et al., 2026). O desenho é observacional e não isola o efeito do sistema de outras mudanças ocorridas no período, ressalva que os próprios autores assumem ao registrar que o estudo documenta melhora da qualidade e da padronização da aquisição, e não da detecção de malformações nem de desfechos perinatais, cuja demonstração depende de estudos que liguem qualidade de imagem a desempenho diagnóstico.

Essa mudança de foco reorganiza a expectativa depositada na tecnologia. Espera-se dela menos o veredito sobre a anomalia e mais a garantia de que o exame foi de fato realizado conforme o protocolo, tarefa que a auditoria manual raramente consegue executar em escala.

RASTREAMENTO CARDÍACO: O TERRITÓRIO DE EVIDÊNCIA MAIS DENSA

O coração fetal concentra a maior parte dos esforços, por reunir alta morbidade, baixa detecção e planos bem definidos. O trabalho que estabeleceu o patamar de referência treinou um conjunto de redes neurais com 107.823 imagens e testou-o em 4.108 exames com prevalência

de 0,9% de cardiopatia, obtendo área sob a curva de 0,99, sensibilidade de 95%, especificidade de 96% e valor preditivo negativo de 100%, com robustez preservada em imagens de outros serviços e de menor qualidade (ARNAOUT R, et al., 2021). Metanálise posterior, reunindo 15 estudos e imagens de 30.121 fetos, encontrou sensibilidade agregada de 0,89 (IC 95%: 0,83-0,93) e especificidade de 0,91 (IC 95%: 0,84-0,95) para a tarefa binária de classificar o coração como normal ou anormal (D'ALBERTI E, et al., 2025).

O detalhe metodológico dessa metanálise vale mais que o resultado agregado. Modelos testados contra um espectro amplo de cardiopatias tiveram sensibilidade menor do que os testados para uma lesão específica, 0,85 contra 0,92, com especificidade equivalente. A heterogeneidade foi substancial, o risco de viés foi classificado como moderado a alto e a aderência aos itens de relato do TRIPOD+AI ficou acima de 70% em apenas 18 dos 51 itens avaliados. O grupo humano de comparação era pequeno, com mediana de seis clínicos, e habitualmente não cegado (D'ALBERTI E, et al., 2025). O desempenho superior ao de operadores menos experientes é consistente; a equivalência a especialistas, muito menos assegurada do que sugerem os números isolados.

Duas estratégias distintas vêm sendo testadas para transpor essa barreira. A primeira substitui a classificação global pela busca de achados morfológicos discretos: avaliação retrospectiva de 877 exames provenientes de 11 centros analisou oito achados associados a cardiopatia grave e registrou sensibilidade de 0,968 para cardiopatia grave e, contra o painel de cardiologistas fetais, sensibilidade de 0,987 e especificidade de 0,977 (ZELOP CM, et al., 2026). A segunda automatiza medidas em vez de emitir juízo categórico. Um modelo de biometria cardíaca aplicado ao exame de 18 a 22 semanas, comparando 73 casos de coartação da aorta a 7.300 controles pareados, atingiu área sob a curva de 0,96, com especificidade de 88,9% para sensibilidade de 90,4% (TAKSØE-VESTER CA, et al., 2024). Como cerca de 60% das coartações isoladas escapam ao diagnóstico pré-natal, a via quantitativa interessa justamente onde o reconhecimento visual falha.

A antecipação do rastreamento para o primeiro trimestre segue a mesma direção. A partir de 108.521 exames cardíacos precoces realizados em múltiplas regiões da China, um modelo interpretável foi construído sobre 8.062 imagens de quatro câmaras com Doppler colorido, com lógica diagnóstica ancorada em padrões de fluxo diastólico e desempenho mantido em validações externas (LEI W, et al., 2026). Uma política de avaliação anatômica

precoce tem efeito populacional mensurável: em coorte nacional inglesa com mais de um milhão de gestações, 32,7% (IC 95%: 31,5%-33,9%) das anomalias maiores foram detectadas antes de 16 semanas, com gradiente crescente conforme a extensão do protocolo adotado pelo serviço (KARIM JN, et al., 2025). O algoritmo atua sobre a acurácia do exame; a política de rastreamento define quantos exames existem para serem acurados.

Modelos voltados a lesões pontuais completam o quadro, ainda que com desenho mais frágil. Sistemas de detecção de comunicação interventricular alcançaram precisão média de 0,879 e mAP de 0,926 em imagens estáticas (LI F, et al., 2024), e comparações entre arquiteturas de detecção reforçaram que a escolha do algoritmo importa menos que a curadoria do conjunto de imagens (YANG Y, et al., 2023). Revisão de escopo dedicada à ecocardiografia fetal automatizada, que examinou 48 trabalhos em detalhe, observou fragmentação semelhante, com predomínio de tarefas isoladas de segmentação e classificação sem encadeamento clínico (HERNANDEZ-CRUZ N, et al., 2025). A Tabela 1 sintetiza o desempenho relatado nos principais estudos por território anatômico e explicita o comparador humano de cada um.

Tabela 1 - Desempenho de sistemas de inteligência artificial na detecção de anomalias fetais por território anatômico

Estudo	Alvo	Conjunto de dados	Desempenho relatado	Comparador humano
Arnaout R, et al. (2021)	Cardiopatias congênitas complexas	107.823 imagens de treino; 4.108 exames de teste, prevalência de 0,9%	AUC 0,99; S 95%; E 96%; VPN 100%	Sensibilidade comparável à de clínicos
Athalye C, et al. (2024)	Cardiopatia grave isolada, rastreamento comunitário	66 casos e 42 controles, imagens armazenadas	S 91%; E 78%	Rastreamento clínico inicial com S 53%; especialistas cegados com S média de 55%
Zelop CM, et al. (2026)	Oito achados morfológicos associados a cardiopatia grave	877 exames de 11 centros	S 0,968 para cardiopatia grave; S 0,987 e E 0,977 contra o painel	Painel de cardiologistas fetais
Taksøe-Vester CA, et al. (2024)	Coartação da aorta por biometria automatizada	73 casos e 7.300 controles pareados	AUC 0,96; E 88,9% com S 90,4%	Não aplicável
Xie HN, et al. (2020)	Encéfalo fetal normal ou anormal	15.372 imagens normais e 14.047 anormais	Acurácia 96,3%; S 96,9%; E 95,9%; AUC 0,989	Não aplicável; lesão localizada com precisão em 61,6% dos casos

Estudo	Alvo	Conjunto de dados	Desempenho relatado	Comparador humano
Lin M, et al. (2022)	Dez padrões de imagem intracraniana	43.890 imagens de 16.297 gestações e 169 vídeos	AUC macro 0,902 na validação externa; 0,969 em varredura em tempo real	Equivalente a ecografistas experientes, com 0,025 s por imagem contra 4,4 s
Zhang Y, et al. (2026)	Fendas orofaciais	45.139 imagens de 9.215 fetos em 22 hospitais	S acima de 93%; E acima de 95%	Equivalente a radiologistas seniores e superior aos juniores
Tang J, et al. (2023)	Trissomias 21, 18 e 13 e doenças genéticas raras	Imagens de face fetal	S de 0,83, 0,92, 0,75 e 0,96; E de 0,92 a 0,95	Comparável a ecografistas experientes
Miguel OX, et al. (2024)	Anomalias renais	969 imagens, sendo 646 controles e 323 casos	AUC 91,3%; S 77,4%; E 87,4%	Não aplicável

Nota: S: sensibilidade; E: especificidade; AUC: área sob a curva; VPN: valor preditivo negativo.
Fonte: elaboração própria a partir dos estudos citados (2026).

SISTEMA NERVOSO CENTRAL, FACE E DEMAIS TERRITÓRIOS

Fora do coração, a evidência se distribui de modo desigual, e a comparação entre territórios revela mais sobre a dificuldade de cada tarefa do que sobre a qualidade dos modelos. O encéfalo fetal, examinado em planos axiais padronizados e com contraste anatômico favorável, produziu os resultados mais altos: classificação de imagens normais e anormais com acurácia de 96,3% e área sob a curva de 0,989, embora a localização da lesão tenha sido considerada precisa por avaliador especialista em apenas 61,6% das imagens anormais (XIE HN, et al., 2020). Acertar a categoria não equivale a apontar o achado, distinção que separa alerta de diagnóstico.

Sistema treinado com 43.890 imagens de 16.297 gestações identificou dez padrões distintos de imagem intracraniana, com área sob a curva macro de 0,902 em validação externa e de 0,969 em varredura em tempo real, desempenho estatisticamente equivalente ao de ecografistas experientes, mas obtido em 0,025 segundo por imagem, contra 4,4 segundos dos especialistas (LIN M, et al., 2022). A diferença temporal, e não a acurácia, é o que torna a triagem contínua concebível durante o exame. Revisão dedicada à neurossonografia assistida mapeia essa expansão, da automatização de medidas à predição de desfechos, e registra a mesma limitação recorrente de validação externa restrita (WEICHERT J e SCHARF JL, 2024).

A face fetal reúne o maior conjunto de treinamento entre os estudos aqui examinados, resultado compatível com a hipótese de que bases amplas e multicêntricas contribuem para robustez, sem que o desenho permita isolar esse fator. Sistema treinado com 45.139 imagens de 9.215 fetos em 22 hospitais diagnosticou fendas orofaciais com sensibilidade superior a 93% e especificidade superior a 95%, equiparando-se a radiologistas seniores e superando os juniores (ZHANG Y, et al., 2026). A mesma região anatômica sustenta abordagem distinta, de rastreamento sindrômico por padrão facial, com sensibilidades de 0,83 para trissomia do 21, 0,92 para trissomia do 18, 0,75 para trissomia do 13 e 0,96 para um conjunto de doenças genéticas raras (TANG J, et al., 2023). O desempenho aparentemente superior nas condições raras convida à cautela: quanto mais raro o achado, menor o número de casos disponíveis e mais instável a estimativa.

O contraste com o trato urinário é instrutivo. Modelo treinado com 969 imagens, das quais 323 de anomalias renais, alcançou área sob a curva de 91,3%, com sensibilidade de 77,4% e especificidade de 87,4% (MIGUEL OX, et al., 2024). O resultado é modesto diante dos anteriores, e o conjunto disponível era muito menor, de quarenta a pouco mais de cem vezes inferior aos empregados nos estudos faciais e cardíacos de maior porte. A relação entre volume de dados e desempenho permanece hipótese plausível, e não conclusão que a comparação entre estudos autorize, já que variam igualmente a anatomia examinada, a complexidade da tarefa, a distribuição dos casos, a arquitetura, a rotulagem e a estratégia de validação. Em serviços com recursos limitados, tarefas biométricas simples podem ser mais viáveis que a classificação de anomalias: fluxo desenvolvido com dados de três instituições etíopes segmentou estruturas cefálicas e classificou micro e macrocefalia por comparação com curvas de referência para a idade gestacional (MENGISTU AK, et al., 2025).

Há ainda uma linha conceitualmente diversa, que não classifica imagens. Sistema de apoio à decisão construído sobre fenótipos perinatais sugere, a cada etapa do exame, qual característica avaliar em seguida para reduzir ao mínimo o conjunto de diagnósticos compatíveis, validado em base de 549 fenótipos de dois centros de referência (STIRNEMANN JJ, et al., 2023). Onde o classificador de imagem responde se algo está errado, o sistema fenotípico organiza a investigação do que já se viu, função particularmente pertinente às doenças raras, em que a limitação do examinador é de memória, não de acuidade visual.

O EXAMINADOR ASSISTIDO: GANHO REAL E RISCO DE DEFERÊNCIA

A pergunta clinicamente relevante não é o quanto o modelo acerta sozinho, mas o que acontece com o desempenho do profissional quando ele recebe a sugestão. A resposta depende de quem recebe e de como a informação é apresentada. Ao incorporar uma representação gráfica explicável à triagem cardíaca, o ganho médio em área sob a curva foi de 0,966 para 0,975 entre especialistas, de 0,829 para 0,890 entre fellows e de 0,616 para 0,748 entre residentes (SAKAI A, et al., 2022). O benefício é inversamente proporcional à experiência prévia, o que sugere que a assistência algorítmica funciona sobretudo como compensação de curva de aprendizado.

Ensaio randomizado cruzado de três vias, com 36 ecografistas de diferentes níveis lendo 709 imagens e vídeos neurossonográficos, elevou a acurácia média de 73% no modo não assistido para 80% com assistência concomitante e 82% quando o sistema era consultado após interpretação inteiramente autônoma, com área sob a curva subindo de 0,85 para 0,90. O modo mais preciso foi também o mais caro em tempo: a mediana de leitura passou de 6 para 11 segundos por caso (LIN M, et al., 2023). Em ensaio randomizado com 32 residentes e 2.720 exames, o grupo treinado com orientação algorítmica necessitou de menos ciclos de treinamento para atingir o padrão de proficiência (p = 0,037), com vantagem na qualidade das imagens obtidas a partir do terceiro ciclo (LEI T, et al., 2024b).

O achado mais desconfortável vem do estudo que examinou a interação em detalhe. Dez clínicos classificaram 500 imagens em quatro condições, totalizando 20 mil classificações, para detecção de defeito do septo atrioventricular. O modelo isolado teve acurácia de 0,798, inferior à dos clínicos sem assistência, de 0,844. A exibição da saída binária do modelo elevou a acurácia conjunta para 0,865. Quando o conselho estava errado, porém, a acurácia caiu de 0,761 para 0,693, e a deterioração foi agravada, e não atenuada, pela exibição do escore de confiança, com acurácia de 0,649, ou dos mapas de ativação (DAY TG, et al., 2024). Informação adicional sobre o funcionamento do modelo aumentou a persuasão sem aumentar o discernimento, resultado que contraria a intuição corrente de que explicabilidade produz uso mais crítico.

Tabela 2 - Efeito da assistência algorítmica sobre o desempenho do examinador

Estudo	Delineamento	Efeito relatado	Ressalva
Sakai A, et al. (2022)	Leitura assistida por representação gráfica explicável, triagem cardíaca	AUC de 0,966 para 0,975 entre especialistas; 0,829 para 0,890 entre fellows; 0,616 para 0,748 entre residentes	Ganho inversamente proporcional à experiência

Estudo	Delineamento	Efeito relatado	Ressalva
Lin M, et al. (2023)	Ensaio randomizado cruzado de três vias, 36 ecografistas, 709 imagens e vídeos	Acurácia de 73% para 80% no modo concomitante e 82% na segunda leitura; AUC de 0,85 para 0,90	Tempo mediano de leitura de 6 s para 11 s no modo mais preciso
Day TG, et al. (2024)	20.000 classificações por 10 clínicos, defeito do septo atrioventricular	Acurácia de 0,844 para 0,865 com saída binária do modelo	Conselho incorreto reduziu a acurácia de 0,761 para 0,693, com piora adicional ao exibir confiança (0,649) ou mapa de ativação
Lei T, et al. (2024b)	Ensaio randomizado com 32 residentes e 2.720 exames	Menos ciclos de treinamento até a proficiência (p = 0,037)	Desfecho de treinamento, não de detecção de anomalia
Li F, et al. (2024)	Diagnóstico assistido de comunicação interventricular	Acurácia de médicos juniores 6,7 pontos percentuais maior; intermediários 2,8 pontos	Estudo retrospectivo com imagens estáticas
Zhang Y, et al. (2026)	Sistema usado como copiloto na leitura de fendas orofaciais	Sensibilidade de radiologistas juniores elevada em mais de 6%	Estudo-piloto com 24 participantes

Nota: AUC: área sob a curva.
Fonte: elaboração própria a partir dos estudos citados (2026).

O conjunto desses resultados, resumido na Tabela 2, autoriza uma leitura prudente. A assistência algorítmica melhora o desempenho médio, com ganho concentrado nos operadores menos experientes, e cobra esse ganho em tempo de exame e em vulnerabilidade ao erro do sistema. Revisão de escopo dos ensaios randomizados de inteligência artificial em prática clínica encontrou 86 estudos, dos quais 81% relataram desfecho primário favorável, com predomínio de centro único e escassa caracterização demográfica das amostras, combinação que recomenda descontar parte do otimismo agregado (HAN R, et al., 2024). Entre os profissionais que executam o exame, a expectativa já é majoritariamente favorável: em levantamento com 151 respondentes, 62% relataram uso corrente de alguma ferramenta, sobretudo biometria automatizada, 87% apoiaram a incorporação e 86% condicionaram esse apoio à manutenção da autoridade clínica final (SMIT RM, et al., 2026). A adesão declarada não é o obstáculo; a governança do erro compartilhado é.

GENERALIZAÇÃO, DESLOCAMENTO DE DOMÍNIO E EQUIDADE

Modelo treinado em um serviço aprende também as características do equipamento, do protocolo e da população daquele serviço. Classificador de planos fetais desenvolvido com dados de 1.792 pacientes na Espanha manteve desempenho em centro dinamarquês, mas exigiu adaptação de domínio para operar em cinco centros africanos que dispunham de apenas 25

pacientes cada; o realinhamento por transferência de aprendizado recuperou a sensibilidade sem sacrificar a precisão (SENDRA-BALCELLS C, et al., 2023). A degradação silenciosa entre serviços é a regra, e não a exceção, o que torna a validação externa condição de segurança, e não refinamento acadêmico.

A trajetória mais consolidada nesse terreno não é a da detecção de anomalias, e a distinção precisa ser mantida com clareza. A estimativa de idade gestacional a partir de varreduras cegas do abdome materno foi validada em 4.695 voluntárias nos Estados Unidos e na Zâmbia, com erro absoluto médio de 3,9 dias contra 4,7 dias da biometria convencional (POKAPRAKARN T, et al., 2022). Integrada a dispositivo portátil de baixo custo e operada por usuários sem qualquer treinamento prévio em ultrassonografia, a ferramenta mostrou equivalência ao padrão do estudo entre 14 e 27 semanas, com erro de 3,2 contra 3,0 dias (STRINGER JSA, et al., 2024). Avaliação subsequente em Chicago e Nairóbi confirmou a transferência para novo ambiente clínico e outra sonda, com erro de 4,1 e 4,3 dias, após adaptação leve com dados locais (WILLIS A, et al., 2026). O que essa linha demonstra é a viabilidade de uma infraestrutura de aquisição simplificada; ela ainda não demonstra que a mesma arquitetura detecte malformações, tarefa de complexidade e de custo de erro incomparavelmente maiores.

A questão da equidade não se resolve por boa intenção. Classificadores de imagem torácica treinados com grandes bases apresentaram subdiagnóstico sistemático em populações historicamente menos assistidas, com efeito ampliado em subgrupos interseccionais (SEYYED-KALANTARI L, et al., 2021). O achado provém de outro domínio de imagem e não se transfere automaticamente à ultrassonografia fetal, mas levanta risco plausível: o falso negativo algorítmico poderia concentrar-se onde a imagem é pior, o operador menos experiente e o acesso ao diagnóstico complementar mais restrito, hipótese que apenas evidência produzida no próprio campo poderá confirmar ou afastar. Sobre o insumo, porém, já existe evidência específica. Exame das bases públicas de ultrassonografia fetal disponíveis para treinar modelos identificou baixa representatividade demográfica, diversidade insuficiente de condições clínicas e heterogeneidade dos equipamentos empregados entre os conjuntos, limitações que podem comprometer robustez, generalização e equidade dos sistemas delas derivados (FIORENTINO MC, et al., 2025). A mitigação exige medida de desempenho estratificada por subgrupo, documentação da composição das bases de treinamento e monitoramento contínuo após a implantação (UEDA D, et al., 2024). Enquanto apenas 22% dos estudos do campo se originarem

de países de renda baixa ou média (GUPTA V, et al., 2026), a promessa de ampliação de acesso permanecerá em grande medida enunciada por quem não a testou nesses cenários.

INCORPORAÇÃO: CUSTO, REGULAÇÃO E RESPONSABILIDADE

A economia da adoção foi estimada em modelo de custo-utilidade que comparou a triagem convencional associada à oximetria de pulso à mesma triagem assistida por aprendizado profundo, ao custo de 145 dólares por exame. Partindo de sensibilidade de 95% contra 58,1%, a incorporação gerou 1.204 anos de vida ajustados por qualidade, com aumento de 22,5 milhões de dólares em custos de rastreamento largamente compensado por 20,4 milhões em redução de custos de tratamento, resultando em 1.720 dólares por ano de vida ajustado por qualidade. O resultado permaneceu favorável para a maior parte das combinações com sensibilidade acima de 80% e especificidade acima de 90% (GINSBERG GM, et al., 2024). A dependência do modelo em relação à especificidade não é detalhe técnico. Com prevalência próxima de 1%, mesmo especificidade de 97,7% produz, por cálculo ilustrativo a partir dos parâmetros relatados, cerca de dois a três falsos positivos para cada verdadeiro positivo, com o correspondente encaminhamento a serviços terciários e o custo emocional imposto às gestantes.

O enquadramento sanitário desses produtos já está definido. No Brasil, a RDC nº 657/2022 disciplina a regularização de software como dispositivo médico e exige, para as classes de risco III e IV, a apresentação de avaliação clínica com associação clínica válida, descrição dos algoritmos empregados, plano de gerenciamento de risco, controles de cibersegurança e declaração de conformidade com as normas IEC 62304, IEC 62366-1 e ISO 14971, além de monitoramento de desempenho no pós-mercado (BRASIL, 2022). A exigência regulatória de descrever a associação clínica válida do algoritmo coincide, em substância, com aquilo que a literatura clínica ainda entrega de modo incompleto.

Na União Europeia, o Regulamento (UE) 2024/1689 passou a ter aplicação geral em 2 de agosto de 2026 e classifica como de risco elevado os sistemas de inteligência artificial que funcionam como componentes de segurança de produtos abrangidos pela legislação de harmonização setorial, categoria que inclui os dispositivos médicos (UNIÃO EUROPEIA, 2024). O calendário foi alterado pelo Regulamento (UE) 2026/1744, em vigor desde 27 de julho de 2026: os requisitos das Seções 1, 2 e 3 do Capítulo III, que reúnem gestão de risco, governança de dados, documentação técnica, transparência, supervisão humana e avaliação de

conformidade, passaram a ser aplicáveis em 2 de dezembro de 2027 aos sistemas classificados como de risco elevado pelo artigo 6(2) e pelo Anexo III, e em 2 de agosto de 2028 aos classificados pelo artigo 6(1) e pelo Anexo I (UNIÃO EUROPEIA, 2026). Um sistema de apoio ao rastreamento fetal incorporado a dispositivo médico não está, portanto, submetido a esse conjunto de obrigações já em 2026, e o dever de registro na base de dados da União, desenhado sobretudo para os sistemas do Anexo III e para os casos do artigo 6(3), não se estende de modo indiscriminado a essa categoria. O adiamento não altera a substância técnica que será exigida: gestão de risco documentada, qualidade e representatividade dos dados de treinamento, supervisão humana efetiva e monitoramento pós-mercado permanecem os eixos sobre os quais a conformidade virá a ser avaliada, e é exatamente sobre eles que a literatura clínica ainda oferece informação incompleta.

É nesse ponto que as diretrizes de relato deixam de ser formalidade. O TRIPOD+AI padronizou o relato de modelos preditivos independentemente do método de estimação (COLLINS GS, et al., 2024); o CONSORT-AI estabeleceu o que deve constar de ensaios com intervenções algorítmicas, incluindo descrição do fluxo de entrada, tratamento de dados de baixa qualidade e análise de erros (LIU X, et al., 2020); e o DECIDE-AI cobriu a fase de avaliação clínica inicial, em pequena escala, na qual se examinam segurança, fatores humanos e integração ao fluxo de trabalho (VASEY B, et al., 2022). A aderência observada permanece baixa (D'ALBERTI E, et al., 2025), e essa lacuna explica por que tantos modelos com métricas excelentes não conseguem sustentar pedido de regularização nem convencer serviços a adotá-los.

Restam questões de responsabilidade que a literatura técnica costuma contornar. Quem responde pelo falso negativo de um sistema regularizado e corretamente operado? Como se registra, no laudo, que o exame foi assistido por software e qual peso teve essa assistência? Editoriais do próprio campo já apontam que o entusiasmo com a detecção de lesões difíceis convive com ausência de ensaios pragmáticos que meçam desfechos perinatais (DRUKKER L, 2022a; DRUKKER L, 2024), e há quem observe que a contribuição real da tecnologia à avaliação do coração fetal ainda se mede mais em promessa que em resultado (QUARELLO E e CORNO E, 2025). No Brasil, a discussão sobre incorporação em ginecologia e obstetrícia tem sido formulada nos mesmos termos, com ênfase na distância entre o entusiasmo de uso e os desafios concretos de implementação (PINHEIRO YT e SILVA RAR, 2024). Registre-se ainda que

dados populacionais consolidados mostram ganho incremental modesto quando se acrescentam exames dirigidos após morfológico normal (CARDINAL MP, et al., 2022), o que sugere que a margem disponível está menos na multiplicação de exames e mais na qualidade daquele que já é universalmente ofertado.

CONCLUSÃO

A evidência disponível sustenta uma afirmação restrita e recusa outra, mais ampla. Sustenta que sistemas de aprendizado profundo aplicados à ultrassonografia obstétrica alcançam, em tarefas circunscritas e com bases suficientemente grandes, acurácia comparável à de especialistas, e que elevam de modo consistente o desempenho de operadores menos experientes, com ganho inversamente proporcional à experiência prévia. Recusa, por ora, a afirmação de que essa acurácia se traduza em aumento demonstrado da detecção pré-natal de malformações em condições reais de rastreamento populacional, porque faltam ensaios prospectivos com desfechos perinatais, a validação externa é escassa e a perda de desempenho associada à baixa qualidade da imagem, observada em estudos específicos, constitui risco relevante ainda não caracterizado para o conjunto dos sistemas.

A contribuição mais plausível no horizonte imediato não está no veredito diagnóstico, e sim na padronização da aquisição, na verificação automática de que todos os planos exigidos foram obtidos com qualidade suficiente e no resgate de casos hoje perdidos por exame incompleto. Trata-se de tecnologia de controle de processo aplicada a um exame em que a variabilidade de execução é uma das principais fragilidades, hipótese que a evidência de mundo real já disponível torna verificável em prazo curto.

Três condições determinarão se esse potencial se realiza. A primeira é a validação externa prospectiva em serviços de rastreamento não terciários, com prevalência real e comparador cegado. A segunda é o desenho da interface, dado que a exibição de mais informação sobre o funcionamento do modelo aumentou a persuasão sem melhorar o discernimento, o que impõe tratar a apresentação da saída algorítmica como variável de segurança, e não de conveniência. A terceira é a distribuição do benefício: sem avaliação estratificada por subgrupo e sem estudos conduzidos em serviços de baixo recurso, o efeito provável é ampliar a distância entre quem já tem acesso a exame de qualidade e quem não tem. A prioridade de pesquisa,

portanto, deslocou-se do desenvolvimento de novos classificadores para a demonstração de que os já existentes funcionam fora do centro que os criou.

REFERÊNCIAS

ARNAOUT, R. et al. An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. Nature Medicine, v. 27, n. 5, p. 882-891, 2021. DOI: 10.1038/s41591-021-01342-5.

ATHALYE, C. et al. Deep-learning model for prenatal congenital heart disease screening generalizes to community setting and outperforms clinical detection. Ultrasound in Obstetrics & Gynecology, v. 63, n. 1, p. 44-52, 2024. DOI: 10.1002/uog.27503.

BOUACHBA, A. et al. Artificial intelligence, radiomics and fetal ultrasound: review of literature and future perspectives. Ultrasound in Obstetrics & Gynecology, v. 65, n. 3, p. 281-291, 2025. DOI: 10.1002/uog.29172.

BRASIL. Agência Nacional de Vigilância Sanitária. Resolução da Diretoria Colegiada RDC nº 657, de 24 de março de 2022. Dispõe sobre a regularização de software como dispositivo médico (Software as a Medical Device - SaMD). Diário Oficial da União: seção 1, Brasília, DF, 30 mar. 2022. Disponível em: https://anvisalegis.datalegis.net/action/ActionDatalegis.php?acao=abrirTextoAto&tipo=RDC&numeroAto=00000657&seqAto=000&valorAno=2022&orgao=RDC/DC/ANVISA/MS. Acesso em: 17 ago. 2026.

CARDINAL, M. P. et al. Incremental detection of severe congenital heart disease by fetal echocardiography following a normal second trimester ultrasound scan in Québec, Canada. Circulation: Cardiovascular Imaging, v. 15, n. 4, e013796, 2022. DOI: 10.1161/CIRCIMAGING.121.013796.

COLLINS, G. S. et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, v. 385, e078378, 2024. DOI: 10.1136/bmj-2023-078378.

D'ALBERTI, E. et al. Artificial intelligence-enabled prenatal ultrasound for the detection of fetal cardiac abnormalities: a systematic review and meta-analysis. EClinicalMedicine, v. 84, 103250, 2025. DOI: 10.1016/j.eclinm.2025.103250.

DAY, T. G. et al. Interaction between clinicians and artificial intelligence to detect fetal atrioventricular septal defects on ultrasound: how can we optimize collaborative performance? Ultrasound in Obstetrics & Gynecology, v. 64, n. 1, p. 28-35, 2024. DOI: 10.1002/uog.27577.

DRUKKER, L. Real-time identification of fetal anomalies on ultrasound using artificial intelligence: what's next? Ultrasound in Obstetrics & Gynecology, v. 59, n. 3, p. 285-287, 2022a. DOI: 10.1002/uog.24869.

DRUKKER, L. The Holy Grail of obstetric ultrasound: can artificial intelligence detect hard-to-identify fetal cardiac anomalies? Ultrasound in Obstetrics & Gynecology, v. 64, n. 1, p. 5-9, 2024. DOI: 10.1002/uog.27703.

DRUKKER, L. et al. Clinical workflow of sonographers performing fetal anomaly ultrasound scans: deep-learning-based analysis. Ultrasound in Obstetrics & Gynecology, v. 60, n. 6, p. 759-765, 2022b. DOI: 10.1002/uog.24975.

FIORENTINO, M. C. et al. Uncovering ethical biases in publicly available fetal ultrasound datasets. npj Digital Medicine, v. 8, art. 355, 2025. DOI: 10.1038/s41746-025-01739-3.

GINSBERG, G. M. et al. Cost-utility analysis of prenatal diagnosis of congenital cardiac diseases using deep learning. Cost Effectiveness and Resource Allocation, v. 22, n. 1, 44, 2024. DOI: 10.1186/s12962-024-00550-3.

GUPTA, V. et al. Emerging applications of artificial intelligence for obstetric ultrasound: a scoping review. International Journal of Gynaecology and Obstetrics, v. 174, n. 1, p. 34-43, 2026. DOI: 10.1002/ijgo.70789.

HAN, R. et al. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. The Lancet Digital Health, v. 6, n. 5, p. e367-e373, 2024. DOI: 10.1016/S2589-7500(24)00047-5.

HERNANDEZ-CRUZ, N. et al. A comprehensive scoping review on machine learning-based fetal echocardiography analysis. Computers in Biology and Medicine, v. 186, 109666, 2025. DOI: 10.1016/j.compbiomed.2025.109666.

HORGAN, R.; NEHME, L.; ABUHAMAD, A. Artificial intelligence in obstetric ultrasound: a scoping review. Prenatal Diagnosis, v. 43, n. 9, p. 1176-1219, 2023. DOI: 10.1002/pd.6411.

KARIM, J. N. et al. Impact of first-trimester ultrasound on early detection of major fetal anomalies: nationwide population-based study of over 1 million pregnancies. PLoS Medicine, v. 22, n. 11, e1004709, 2025. DOI: 10.1371/journal.pmed.1004709.

LEI, T. et al. Development and validation of an artificial intelligence assisted prenatal ultrasonography screening system for trainees. International Journal of Gynaecology and Obstetrics, v. 165, n. 1, p. 306-317, 2024a. DOI: 10.1002/ijgo.15167.

LEI, T. et al. Enhancing trainee performance in obstetric ultrasound through an artificial intelligence system: randomized controlled trial. Ultrasound in Obstetrics & Gynecology, v. 64, n. 4, p. 453-462, 2024b. DOI: 10.1002/uog.29101.

LEI, W. et al. An interpretable deep learning model for first-trimester fetal cardiac screening. npj Digital Medicine, v. 9, art. 43, 2026. Publicação eletrônica antecipada em 8 dez. 2025. DOI: 10.1038/s41746-025-02217-6.

LEIROZ, R. et al. Accuracy of the mid-trimester ultrasound scan in the detection of fetal congenital anomalies in a reference center in Northeastern Brazil. Journal of Gynecology Obstetrics and Human Reproduction, v. 50, n. 10, 102225, 2021. DOI: 10.1016/j.jogoh.2021.102225.

LI, F. et al. Application of artificial intelligence in VSD prenatal diagnosis from fetal heart ultrasound images. BMC Pregnancy and Childbirth, v. 24, n. 1, 758, 2024. DOI: 10.1186/s12884-024-06916-y.

LIN, M. et al. Use of real-time artificial intelligence in detection of abnormal image patterns in standard sonographic reference planes in screening for fetal intracranial malformations. Ultrasound in Obstetrics & Gynecology, v. 59, n. 3, p. 304-316, 2022. DOI: 10.1002/uog.24843.

LIN, M. et al. Deep learning system improved detection efficacy of fetal intracranial malformations in a randomized controlled trial. npj Digital Medicine, v. 6, n. 1, 191, 2023. DOI: 10.1038/s41746-023-00932-6.

LIU, X. et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nature Medicine, v. 26, n. 9, p. 1364-1374, 2020. DOI: 10.1038/s41591-020-1034-x.

MENGISTU, A. K. et al. Detecting microcephaly and macrocephaly from ultrasound images using artificial intelligence. BMC Medical Imaging, v. 25, n. 1, 183, 2025. DOI: 10.1186/s12880-025-01709-x.

MIGUEL, O. X. et al. Deep learning prediction of renal anomalies for prenatal ultrasound diagnosis. Scientific Reports, v. 14, n. 1, 9013, 2024. DOI: 10.1038/s41598-024-59248-4.

PINHEIRO, D. O. et al. Accuracy of prenatal diagnosis of congenital cardiac malformations. Revista Brasileira de Ginecologia e Obstetrícia, v. 41, n. 1, p. 11-16, 2019. DOI: 10.1055/s-0038-1676058.

PINHEIRO, Y. T.; SILVA, R. A. R. da. Artificial intelligence in gynecology and obstetrics: from the enthusiasm of use in practice to the challenges of implementation. Revista Brasileira de Ginecologia e Obstetrícia, v. 46, e-rbg041, 2024. DOI: 10.61622/rbgo/2024rbg041.

POKAPRAKARN, T. et al. AI estimation of gestational age from blind ultrasound sweeps in low-resource settings. NEJM Evidence, v. 1, n. 5, 2022. DOI: 10.1056/EVIDoa2100058.

QUARELLO, E.; CORNO, E. Thoughts on the contribution of artificial intelligence (AI) to assessment of the fetal heart: a true scientific odyssey. Ultrasound in Obstetrics & Gynecology, v. 65, n. 3, p. 292-294, 2025. DOI: 10.1002/uog.29185.

RAMIREZ ZEGARRA, R.; GHI, T. Use of artificial intelligence and deep learning in fetal ultrasound imaging. Ultrasound in Obstetrics & Gynecology, v. 62, n. 2, p. 185-194, 2023. DOI: 10.1002/uog.26130.

SAKAI, A. et al. Medical professional enhancement using explainable artificial intelligence in fetal cardiac ultrasound screening. Biomedicines, v. 10, n. 3, 551, 2022. DOI: 10.3390/biomedicines10030551.

SALOMON, L. J. et al. ISUOG practice guidelines (updated): performance of the routine mid-trimester fetal ultrasound scan. Ultrasound in Obstetrics & Gynecology, v. 59, n. 6, p. 840-856, 2022. DOI: 10.1002/uog.24888.

SENDRA-BALCELLS, C. et al. Generalisability of fetal ultrasound deep learning models to low-resource imaging settings in five African countries. Scientific Reports, v. 13, n. 1, 2728, 2023. DOI: 10.1038/s41598-023-29490-3.

SEYYED-KALANTARI, L. et al. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nature Medicine, v. 27, n. 12, p. 2176-2182, 2021. DOI: 10.1038/s41591-021-01595-0.

SMIT, R. M. et al. Artificial intelligence as an add-on instrument in fetal ultrasound: sonographers' and obstetricians' expectations. Prenatal Diagnosis, 2026. Publicação eletrônica antecipada. DOI: 10.1002/pd.70133.

STIRNEMANN, J. J. et al. Development and clinical validation of real-time artificial intelligence diagnostic companion for fetal ultrasound examination. Ultrasound in Obstetrics & Gynecology, v. 62, n. 3, p. 353-360, 2023. DOI: 10.1002/uog.26242.

STRINGER, J. S. A. et al. Diagnostic accuracy of an integrated AI tool to estimate gestational age from blind ultrasound sweeps. JAMA, v. 332, n. 8, p. 649-657, 2024. DOI: 10.1001/jama.2024.10770.

TAKSØE-VESTER, C. A. et al. Role of artificial-intelligence-assisted automated cardiac biometrics in prenatal screening for coarctation of aorta. Ultrasound in Obstetrics & Gynecology, v. 64, n. 1, p. 36-43, 2024. DOI: 10.1002/uog.27608.

TANG, J. et al. A deep-learning-based method can detect both common and rare genetic disorders in fetal ultrasound. Biomedicines, v. 11, n. 6, 1756, 2023. DOI: 10.3390/biomedicines11061756.

UEDA, D. et al. Fairness of artificial intelligence in healthcare: review and recommendations. Japanese Journal of Radiology, v. 42, n. 1, p. 3-15, 2024. DOI: 10.1007/s11604-023-01474-3.

UNIÃO EUROPEIA. Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial (Regulamento da Inteligência Artificial). Jornal Oficial da União Europeia, L, 2024/1689, 12 jul. 2024. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689. Acesso em: 17 ago. 2026.

UNIÃO EUROPEIA. Regulamento (UE) 2026/1744 do Parlamento Europeu e do Conselho, de 8 de julho de 2026, que altera os Regulamentos (UE) 2024/1689, (UE) 2018/1139 e (UE) 2023/1230 no que diz respeito à simplificação da aplicação das regras harmonizadas em matéria

de inteligência artificial (Regulamento Omnibus Digital em matéria de IA). Jornal Oficial da União Europeia, L, 2026/1744, 24 jul. 2026. Disponível em: https://eur-lex.europa.eu/eli/reg/2026/1744/oj. Acesso em: 17 ago. 2026.

VAN VELZEN, C. L. et al. Systematic review and meta-analysis of the performance of second-trimester screening for prenatal detection of congenital heart defects. International Journal of Gynaecology and Obstetrics, v. 140, n. 2, p. 137-145, 2018. DOI: 10.1002/ijgo.12373.

VASEY, B. et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nature Medicine, v. 28, n. 5, p. 924-933, 2022. DOI: 10.1038/s41591-022-01772-9.

WEICHERT, J.; SCHARF, J. L. Advancements in artificial intelligence for fetal neurosonography: a comprehensive review. Journal of Clinical Medicine, v. 13, n. 18, 5626, 2024. DOI: 10.3390/jcm13185626.

WILLIS, A. et al. Generalization of AI-based gestational age assessment using blind sweep ultrasonography. JAMA Network Open, v. 9, n. 7, e2622484, 2026. DOI: 10.1001/jamanetworkopen.2026.22484.

XIE, H. N. et al. Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal. Ultrasound in Obstetrics & Gynecology, v. 56, n. 4, p. 579-587, 2020. DOI: 10.1002/uog.21967.

YANG, Y. et al. Classification of normal and abnormal fetal heart ultrasound images and identification of ventricular septal defects based on deep learning. Journal of Perinatal Medicine, v. 51, n. 8, p. 1052-1058, 2023. DOI: 10.1515/jpm-2023-0041.

ZELOP, C. M. et al. Artificial intelligence for the detection of fetal ultrasound findings concerning for major congenital heart defects. Obstetrics and Gynecology, v. 147, n. 1, p. 97-107, 2026. DOI: 10.1097/AOG.0000000000006027.

ZHANG, Y. et al. Artificial intelligence for detecting fetal orofacial clefts and advancing medical education. Nature Communications, v. 17, n. 1, 7637, 2026. DOI: 10.1038/s41467-026-74119-4.

ZHAO, J. et al. AI-based quality control was associated with improved fetal ultrasound image quality in low-resource settings: a real-world multicenter study from West China. BMC Medicine, v. 24, art. 235, 2026. DOI: 10.1186/s12916-026-04768-1.