



Artificial Intelligence for Detecting Fetal Genetic Variants Across the Prenatal Testing Pathway: A Critical Narrative Review

**Stefan Bittmann^{a,b,++*}, Elisabeth Luchter^a
and Elena Moschüring-Alieva^a**

^a *Department of Pediatrics, Ped Mind Institute, Hindenburgring 4, D-48599, Gronau, Germany.*

^b *School of Medicine, Shangluo Vocational and Technical College, Shangluo, 726000, Shaanxi, China.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: <https://doi.org/10.9734/cjast/2026/v45i104782>

Open Peer Review History:

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: <https://pr.sdiarticle5.com/review-history/167077>

Review Article

Received: 09/08/2026

Accepted: 28/09/2026

Published: 01/10/2026

Abstract

Background: Prenatal genetic testing now spans screening of cell-free DNA in maternal plasma, sonographic phenotyping, cytogenetic analysis and exome or genome sequencing of fetal samples. Artificial intelligence, including machine learning and deep learning, has been applied at each point, yet the literature is fragmented across laboratory genomics, imaging science and clinical genetics, and its evidential maturity has rarely been compared across domains.

Purpose and scope: This critical narrative review examines how artificial intelligence has been used to detect or interpret fetal aneuploidy, copy number variants, single-nucleotide variants and syndromic phenotypes, and evaluates the strength, validity and limitations of that evidence. Embryo selection in assisted reproduction and prediction of non-genetic obstetric outcomes were excluded.

Approach: Literature from January 1995 to July 2026 was identified through bibliographic databases, a trial registry and citation searching, and was appraised against reporting and validation principles for diagnostic artificial intelligence.

⁺⁺ *Honorary Professor (Hon. Professor);*

^{*} *Corresponding author: E-mail: stefanbittmann@gmx.de;*

Principal findings: Evidence is strongest for supportive analytical tasks, notably fetal DNA fraction estimation, automated chromosome classification and prioritisation of candidate variants, where algorithms operate under human oversight against measurable reference standards. Machine-learning refinement of aneuploidy calls from cell-free DNA has improved positive predictive value in a large retrospective series, but the incremental benefit over established screening is bounded by placental mosaicism and maternal confounders rather than by computation. Genome-wide non-invasive genotyping of monogenic variants has been demonstrated only in small family cohorts. Image-based prediction of chromosomal or syndromic disease from ultrasound remains developmental, relying on case-control sampling, single-centre data and scarce external validation. Prenatal sequencing interpretation is constrained by incomplete fetal phenotypes, ancestry bias in reference data and unverified large language model outputs.

Unresolved questions: Prospective, multicentre evaluations with appropriate comparators, fetal rather than placental reference standards and assessment of counselling and decision outcomes are largely absent.

Conclusion: Artificial intelligence is best regarded as an adjunct that improves efficiency and consistency within existing testing pathways rather than as an autonomous detector of fetal genetic disease. Safe expansion depends on transparent reporting, equitable datasets, calibrated uncertainty and governance that preserves informed reproductive choice.

Keywords: *Cell-free DNA; non-invasive prenatal testing; deep learning; fetal aneuploidy; prenatal exome sequencing; variant interpretation; automated karyotyping; algorithmic bias.*

1. Introduction

Syndromic genetic conditions have been estimated to affect, in aggregate, about 8% of the population (Gurovich *et al.*, 2019), and their prenatal identification shapes counselling, pregnancy management, delivery planning and reproductive choice (Van den Veyver *et al.*, 2022). The diagnostic task is heterogeneous. Whole-chromosome aneuploidies such as trisomy 21, trisomy 18 and trisomy 13 alter gene dosage across an entire chromosome; copy number variants (CNVs) involve gains or losses of smaller genomic segments; single-nucleotide variants (SNVs) and small insertions or deletions disrupt individual genes; and mosaicism distributes any of these changes unevenly across fetal and placental cell lineages. Each class demands different specimens, laboratory methods and interpretive frameworks, which is why contemporary prenatal genetics relies on a sequence of tests rather than on a single assay (Sparks & Chitty, 2026; Van den Veyver *et al.*, 2022).

The terminology used in this field requires clarification. The title of this review refers to fetal mutations in the colloquial sense, but professional standards recommend the neutral term variant, qualified by a pathogenicity classification, because many DNA changes identified in a fetus are inherited, benign or of uncertain significance (Richards *et al.*, 2015; Riggs *et al.*, 2020). The present article therefore uses fetal genetic variant as the umbrella term, reserving mutation for historical usage or for *de novo* events described as such by the original authors. Artificial intelligence (AI) is used to denote computational systems that learn decision rules from data, encompassing classical machine learning (ML), such as support vector machines and random forests, and deep learning (DL), which uses multilayer neural networks, including convolutional neural networks (CNNs) for images and sequence data. Large language models (LLMs) are treated as a distinct and more recent subclass.

The computational character of prenatal genetic testing is not new. Artificial neural networks were applied to maternal serum markers for the detection of fetal chromosomal disorders in the mid-1990s (Beksaç *et al.*, 1995), well before the widespread adoption of DL. The subsequent discovery of fetal DNA in maternal plasma (Lo *et al.*, 1997) and the demonstration that massively parallel sequencing of plasma DNA could reveal fetal trisomy 21 (Chiu *et al.*, 2008) turned prenatal screening into a problem of statistical inference over millions of short DNA fragments. Shortly afterwards, deep sequencing of maternal plasma was shown to contain a representation of the entire fetal genome, allowing inference of fetal genotypes when parental haplotypes were known (Fan *et al.*, 2012; Kitzman *et al.*, 2012; Lo *et al.*, 2010). Non-invasive prenatal testing (NIPT) based on cell-free DNA (cfDNA) now detects more than 99% of trisomy 21 at a false-positive rate of 0.04% in pooled analyses (Gil *et al.*, 2017), and a large prospective comparison in a routine population reported a positive predictive value (PPV) of 80.9% for cfDNA testing against 3.4% for conventional combined screening (Norton *et al.*, 2015). Professional guidance now recommends cfDNA screening for common trisomies in all singleton and twin pregnancies (Dungan *et al.*, 2023). Recent synthesis indicates that non-invasive prenatal screening is extending beyond common fetal aneuploidies toward broader analysis of cfDNA-derived genetic and molecular signals, increasing the interpretive and computational demands of prenatal screening (Stanley *et al.*, 2025).

In parallel, diagnostic testing of invasive samples has progressed from karyotyping and chromosomal microarray analysis to exome sequencing (ES) and genome sequencing. Prospective cohorts in the United Kingdom and the United States identified diagnostic variants in approximately 8.5% and 10% of fetuses with structural anomalies after normal karyotype and microarray results (Lord *et al.*, 2019; Petrovski *et al.*, 2019), and a subsequent meta-analysis estimated a pooled incremental yield of 31%, rising to 42% in cases preselected for a likely monogenic cause (Mellis *et al.*, 2022). These gains have come at the cost of analytical complexity. Each fetal exome yields tens of thousands of variants that must be filtered, prioritised and classified against a phenotype that is incomplete, evolving with gestation and often limited to ultrasound findings (Shear *et al.*, 2025). Recent review of fetal genome sequencing similarly emphasizes that expanded variant detection must be accompanied by cautious interpretation of evolving prenatal phenotypes and bioinformatic pipeline limitations (Basel-Salmon & Brabbing-Goldstein, 2024).

AI has entered this landscape from several directions. In cfDNA analysis it has been used to estimate the fetal DNA fraction, refine aneuploidy calls and infer fetal genotypes (Tsui *et al.*, 2025). In imaging it has been trained to recognise sonographic markers, facial morphology and structural anomalies that raise the probability of a genetic diagnosis (Horgan *et al.*, 2023; Umans *et al.*, 2024). In cytogenetics it now segments and classifies metaphase chromosomes in commercial software used by clinical laboratories

(Rosenblum *et al.*, 2025), and in genomic medicine it supports variant calling, pathogenicity prediction and phenotype-driven prioritisation (Cheng *et al.*, 2023; Poplin *et al.*, 2018; Vestito *et al.*, 2024). Enthusiasm has been considerable. Several primary studies describe their models as accurate, expert-level or ready to facilitate universal screening, and a review of AI for prenatal chromosome analysis concluded that ML and DL had substantially improved detection while acknowledging small samples and uncertain generalisability (Boddupally & Rani Thuraka, 2024). Within clinical genetics more broadly, AI is increasingly being considered for diagnostic interpretation, variant prioritisation and clinical support, reinforcing its relevance to prenatal genomic workflows while retaining the need for expert review (Duong & Solomon, 2025). A recent review of fetal ultrasound AI likewise describes expanding use in automated anatomical assessment and anomaly detection while underscoring the importance of dataset quality and validation for clinical translation (Bouachba *et al.*, 2025).

Existing reviews, however, leave important questions unresolved. Reviews of AI in cfDNA diagnostics have concentrated on methodological innovation across oncology and obstetrics without a dedicated appraisal of fetal genetic endpoints (Tsui *et al.*, 2025). Reviews of AI in obstetric ultrasound catalogue a wide range of tasks, most of them biometric or anatomical, and report that few algorithms have been externally validated (Horgan *et al.*, 2023; Umans *et al.*, 2024). A meta-analysis of AI for fetal cardiac anomalies found pooled sensitivity and specificity near 0.9 but a moderate-to-high risk of bias and poor adherence to reporting standards (D'Alberti *et al.*, 2025). Reviews of AI-guided karyotyping describe laboratory experience rather than diagnostic accuracy for abnormal fetal karyotypes (Rosenblum *et al.*, 2025), and ethical analyses of genomic AI in early life have not been linked to the empirical evidence on performance (Coghlan *et al.*, 2024). No review identified for the present work has compared these domains within a single framework that distinguishes the variant class being detected, the biological source of the signal, the maturity of validation and the clinical role of the algorithm. This omission matters because claims about AI-based detection of fetal genetic disease are frequently transferred between domains whose error structures are fundamentally different.

1.1 Scope, Research Gap and Objectives

This review addresses the use of AI for detecting, estimating or interpreting fetal genetic variation in human pregnancy. Four application domains are included: analysis of cfDNA in maternal plasma; image-based inference of genetic risk from ultrasound and related fetal imaging; cytogenetic analysis of prenatal samples; and interpretation of fetal sequencing data, including CNV classification and language-model support. The variant classes considered are whole-chromosome aneuploidy, subchromosomal CNVs, SNVs and small insertions or deletions, and mosaicism where it affects test interpretation. Algorithms that predict non-genetic obstetric outcomes such as pre-eclampsia, preterm birth or fetal growth restriction are excluded, as are AI tools for embryo selection and preimplantation genetic testing, because the embryo, the specimen and the decision context differ from those of an established pregnancy. Postnatal facial analysis and general rare-disease genomics are considered only where they inform prenatal translation.

The specific gap addressed is the absence of an integrated, critically appraised synthesis that relates AI performance claims in each prenatal domain to the biological ceiling imposed by the signal source, to the adequacy of the reference standard and to the stage of validation reached. The objectives are to define the principal AI tasks across the prenatal genetic pathway; to evaluate the strength and consistency of evidence for each task; to identify recurring methodological weaknesses and explain discrepancies among studies; to examine equity, ethical and governance implications that arise specifically in the prenatal setting; and to set out prioritised research needs. The intended contribution is an evidence-calibrated account that separates demonstrated analytical gains from preliminary or speculative claims of autonomous fetal genetic detection.

2. Methods for Literature Selection

2.1 Review Design and Rationale

A critical narrative review design was selected because the literature on AI for fetal genetic detection is methodologically heterogeneous. It combines diagnostic accuracy studies, algorithm development papers from computer science, laboratory validation reports, cohort studies of sequencing yield, reporting guidelines and ethical analyses. These sources differ in their units of analysis (the chromosome, the image, the fetus or the family), in their reference standards and in their outcome metrics, so that statistical pooling across domains would have been neither meaningful nor defensible. A systematic review or meta-analysis would have been appropriate for a single, well-defined index test and target condition, and such syntheses already exist for some imaging applications (D'Alberti *et al.*, 2025; Umans *et al.*, 2024). The present question, by contrast, concerns how evidence of different kinds and different maturity relates across the prenatal testing pathway, a question for which interpretive synthesis is better suited (Greenhalgh *et al.*, 2018). Narrative design does not remove the obligation of transparency. The review was therefore planned and reported with reference to the six quality domains of the Scale for the Assessment of Narrative Review Articles, namely justification of importance, statement of aims, description of the literature search, referencing, presentation of levels of evidence and presentation of relevant endpoint data (Baethge *et al.*, 2019).

2.2 Information Sources

Bibliographic searching used PubMed, Europe PMC (including its PubMed Central full-text collection), Crossref, Semantic Scholar, Google Scholar and the Directory of Open Access Journals (DOAJ). These sources were chosen to capture biomedical, genomic and computational literature, including engineering and informatics journals that publish algorithm development studies. The ClinicalTrials.gov registry was searched to identify registered prospective evaluations of AI in prenatal genetic testing. Institutional sources comprised the official publication pages of the World Health Organization and position statements or guidelines of professional societies in prenatal diagnosis and medical genetics published in peer-reviewed journals. Database searching was

supplemented by backward citation searching of included reviews and key primary studies, forward citation searching of seminal papers, related-article searching and targeted checks for corrections, errata, expressions of concern and retractions.

2.3 Search Strategy and Search Period

Search concepts were constructed around four blocks: AI and computational methods; the prenatal or fetal context; the genetic target, including aneuploidy, CNVs, monogenic variants and syndromes; and the data modality, including cfDNA, ultrasound, karyotype and sequencing. Additional terms addressed validation, bias, ethics and implementation. Controlled vocabulary (Medical Subject Headings in PubMed) was combined with free-text synonyms using Boolean operators, and strings were adapted to the syntax of each source. Table 1 summarises the concept blocks and representative strings. The search period extended from 1 January 1995, corresponding to the earliest identified application of neural networks to prenatal genetic disorder detection (Beksaç *et al.*, 1995), to 26 July 2026. Foundational publications on cfDNA biology and genome-wide fetal sequencing that informed the conceptual framework were retained irrespective of their relation to the search start.

Table 1. Search concept blocks and representative Boolean strings used to identify literature on artificial intelligence for the detection of fetal genetic variants

Concept block	Principal controlled and free-text terms	Representative Boolean string (PubMed syntax)	Supporting references
1. AI and computational methods	Artificial intelligence [MeSH]; machine learning [MeSH]; deep learning; neural network; convolutional; random forest; support vector machine; large language model	("artificial intelligence"[MeSH] OR "machine learning"[MeSH] OR "deep learning" OR "neural network*" OR "random forest" OR "support vector machine*" OR "large language model*")	Collins <i>et al.</i> , 2024; Tsui <i>et al.</i> , 2025
2. Prenatal context	Prenatal diagnosis [MeSH]; fetus [MeSH]; fetal; foetal; prenatal; antenatal	(prenatal OR antenatal OR fetal OR foetal OR fetus OR "prenatal diagnosis"[MeSH])	Horgan <i>et al.</i> , 2023; Van den Veyver <i>et al.</i> , 2022
3. Genetic target	Aneuploidy [MeSH]; trisomy; copy number variation; microdeletion; monogenic; single-nucleotide variant; mutation; mosaicism; syndrome	(aneuploid* OR trisomy OR "copy number" OR microdeletion* OR monogenic OR "single nucleotide variant*" OR mutation* OR mosaic* OR syndrom*)	Grati <i>et al.</i> , 2020; Richards <i>et al.</i> , 2015; Riggs <i>et al.</i> , 2020
4a. Cell-free DNA modality	Cell-free nucleic acids [MeSH]; cell-free DNA; cfDNA; non-invasive prenatal testing; NIPT; NIPS; fetal fraction; fragmentomics	("cell-free DNA" OR cfDNA OR "noninvasive prenatal" OR "non-invasive prenatal" OR NIPT OR NIPS OR "fetal fraction" OR fragment*)	Peng & Jiang, 2017; Tsui <i>et al.</i> , 2025
4b. Imaging modality	Ultrasonography, prenatal [MeSH]; nuchal translucency; facial profile; fetal echocardiography; fetal MRI	(ultrasound OR ultrasonograph* OR sonograph* OR "nuchal translucency" OR "fetal echocardiograph*" OR "fetal MRI")	Meshaka <i>et al.</i> , 2023; Umans <i>et al.</i> , 2024
4c. Cytogenetic modality	Karyotyping [MeSH]; karyotype; metaphase; chromosome classification; chromosome segmentation; amniotic fluid	(karyotyp* OR metaphase OR "chromosome classification" OR "chromosome segmentation") AND (prenatal OR amniotic OR clinical)	Rosenblum <i>et al.</i> , 2025
4d. Sequencing modality	Exome sequencing [MeSH]; whole genome sequencing [MeSH]; variant interpretation; variant prioritisation; Human Phenotype Ontology; phenotype-driven	(exome OR "genome sequencing") AND ("variant interpretation" OR priorit* OR "phenotype-driven" OR "Human Phenotype Ontology" OR Exomiser)	Duyzend <i>et al.</i> , 2024; Shear <i>et al.</i> , 2025
5. Evaluation and implementation	Validation; external validation; generalisability; bias; equity; ethics; reporting guideline; implementation	("external validation" OR generalis* OR generaliz* OR bias OR equity OR ethic* OR implementation OR "reporting guideline")	Coghlan <i>et al.</i> , 2024; Sounderajah <i>et al.</i> , 2025

Note. Core searches combined block 1 AND block 2 AND block 3 AND at least one modality block (4a to 4d); block 5 was added in supplementary searches on validation, equity and governance. Strings were adapted to the syntax of each information source. AI = artificial intelligence; cfDNA = cell-free DNA; MeSH = Medical Subject Headings; MRI = magnetic resonance imaging; NIPS = non-invasive prenatal screening; NIPT = non-invasive prenatal testing

2.4 Eligibility Criteria

Eligible publications were peer-reviewed original studies, systematic or scoping reviews, meta-analyses, consensus statements, reporting guidelines and ethical or conceptual analyses that addressed AI applied to the detection, estimation or interpretation of fetal genetic variation in human pregnancy, or that provided essential context on the reference tests, signal biology or evaluation standards against which such AI must be judged. Primary AI studies were required to describe the data source, the target condition, the reference standard and at least one quantitative performance measure. Studies were excluded when they addressed non-genetic obstetric prediction, embryo selection or preimplantation testing, non-human data, or cancer liquid biopsy without transferable prenatal relevance. Conference abstracts, theses, preprints, commercial materials and publications whose bibliographic details could not be verified were excluded. Publications in English were eligible, and non-English publications with English abstracts were screened but were included only when the full bibliographic record and a verifiable citation could be confirmed. No geographical restriction was applied.

2.5 Screening, Duplicate Handling and Study Selection

Titles and abstracts retrieved from each source were screened for relevance to the review scope, and potentially eligible records were assessed in full text when available. Records retrieved from more than one source were identified as duplicates by Digital Object Identifier, PubMed identifier and title matching, and a single record was retained. Bibliographic details of every included

publication were verified against the publisher record, PubMed or Crossref, and the resolution of each Digital Object Identifier to the cited article was confirmed. Publications that could not be accessed sufficiently to verify the cited claim were not used to support specific statements. Consistent with narrative methodology, selection prioritised relevance, methodological adequacy, representativeness of each application domain and the inclusion of contradictory or negative findings, rather than completeness of enumeration. Record counts are therefore not reported, and no flow diagram is presented because a formal systematic selection process was not undertaken.

2.6 Critical Appraisal and Evidence Synthesis

Primary AI studies were appraised using domains drawn from reporting and evaluation guidance for prediction models and diagnostic AI, including the transparent reporting guidance for prediction models that use ML (Collins *et al.*, 2024), the reporting guideline for AI-centred diagnostic accuracy studies (Sounderajah *et al.*, 2025), guidance on early clinical evaluation of AI decision support (Vasey *et al.*, 2022) and guidance on trials of AI interventions (X. Liu *et al.*, 2020). The domains considered were the sampling frame (consecutive cohort or case-control enrichment), sample size and number of positive cases, separation of training and test data at the level of patient or pregnancy, the nature of the reference standard, external or temporal validation, comparator choice, calibration and handling of uncertain outputs, and reporting of subgroup performance. Formal risk-of-bias scores were not assigned because many algorithm development studies did not report the information required for a recognised tool to be applied reliably. Each study was instead assigned to one of four evidence-maturity levels: technical development with internal validation only; external or temporal validation on independent data; prospective or real-world clinical evaluation; and evaluation of clinical outcomes or implementation. Themes were developed iteratively by grouping studies according to the variant class targeted and the signal source used, and discrepancies between studies were interpreted with reference to differences in population, prevalence, reference standard, comparator and analytical unit.

Table 1 presents the search concepts and representative Boolean strings. The strings illustrate the logic of the search rather than a fixed query, because syntax was adapted to each source, and the supporting references indicate publications whose terminology informed the construction of each concept block.

3. What Artificial Intelligence Is Asked to Detect: Signals, Tasks and Biological Ceilings

3.1 Variant Classes and the Sources of Fetal Genetic Signal

Any evaluation of AI for fetal genetic detection must begin with the signal on which the algorithm operates, because the signal determines the upper limit of achievable accuracy regardless of computational sophistication. Four signal sources can be distinguished. The first is cfDNA in maternal plasma, in which fetal-derived fragments form a minority population among predominantly maternal fragments. The second is the fetal phenotype captured by imaging, which may reflect a genetic disorder only indirectly and incompletely. The third is the chromosome complement of cultured or uncultured fetal cells obtained by amniocentesis or chorionic villus sampling, visualised by banding. The fourth is the DNA sequence of a fetal sample, generated by microarray, exome or genome sequencing and interpreted against population, clinical and phenotypic databases.

The cfDNA signal carries a particular biological constraint that is frequently underemphasised in AI studies. The so-called fetal fraction of cfDNA derives predominantly from the placenta, specifically from trophoblast, rather than from the fetus itself (Merhala *et al.*, 2026). Confined placental mosaicism (CPM) can therefore produce an abnormal cfDNA result in a chromosomally normal fetus, and genome-wide cfDNA screening frequently identifies rare autosomal trisomies that are largely attributable to CPM, many of which carry little risk to the fetus (Grati *et al.*, 2020). Maternal factors impose further ceilings. Maternal CNVs, uterine fibroids and occult maternal malignancy can generate copy-number patterns in plasma that mimic or obscure fetal findings (Merhala *et al.*, 2026; Turriff *et al.*, 2024). In a case series drawn from 125,426 clinical NIPT samples, maternal cancers clustered among the rare results showing more than one aneuploidy (Bianchi *et al.*, 2015), and a later prospective study found occult cancer in 48.6% of pregnant or postpartum individuals referred after unusual or non-reportable cfDNA results, with a pattern of gains and losses across three or more chromosomes being strongly associated with malignancy (Turriff *et al.*, 2024). A recent review has characterised CNV detection by NIPT as probabilistic inference on a composite maternal-placental signal rather than a direct measurement of fetal copy number (Merhala *et al.*, 2026). These observations imply that any algorithm trained to predict fetal karyotype from plasma DNA learns, in part, to predict placental and maternal states, and that its apparent errors may reflect genuine biological discordance rather than computational failure.

Imaging signals present a different constraint. Sonographic markers such as increased nuchal translucency (NT), absent nasal bone or abnormal facial angles are associated with aneuploidy but lack specificity, and many monogenic disorders have no characteristic prenatal phenotype or develop their distinguishing features late in gestation (Shear *et al.*, 2025). Even among fetuses with ultrasound-detected structural anomalies, the probability of a molecular diagnosis varies widely by organ system, from a few per cent for isolated increased NT to more than half for isolated skeletal anomalies (Mellis *et al.*, 2022). An imaging algorithm can therefore at best identify a raised probability of genetic disease and cannot establish its molecular basis.

Cytogenetic and sequencing signals are closer to the fetal genotype but are limited by resolution and interpretation. Conventional G-banded karyotyping visualises the whole genome at single-cell resolution and detects aneuploidy, balanced and unbalanced rearrangements and mosaicism (Zhu *et al.*, 2025), but submicroscopic CNVs require copy-number analysis by microarray or sequencing (Riggs *et al.*, 2020). Sequencing detects variants at base-pair resolution but generates an interpretive burden, and its diagnostic yield depends heavily on how the fetal phenotype is encoded and how candidate variants are prioritised (Fu *et al.*, 2022; Mellis *et al.*, 2022).

Table 2. Taxonomy of artificial intelligence applications across the prenatal genetic testing pathway, by signal source, variant class, analytical task, clinical role and dominant performance constraint

Application	Signal source	Variant class targeted	AI task	Clinical role	Dominant performance constraint	Supporting references
Fetal fraction estimation	cfDNA read counts, fragment size and genomic distribution	Indirect; quality parameter for all classes	Estimation (multivariate regression, neural networks)	Quality control; precondition for calling	No gold standard in female pregnancies; inter-method variability	Becking <i>et al.</i> , 2023; Kim <i>et al.</i> , 2015; Yuan <i>et al.</i> , 2020
Aneuploidy calling	cfDNA read counts and fragment distances	Trisomies 21, 18 and 13	Classification (SVM, CNN)	Screening	Placental mosaicism; high baseline performance limits gains	Gil <i>et al.</i> , 2017; Lee <i>et al.</i> , 2022; J. Yang <i>et al.</i> , 2018
Subchromosomal and genome-wide cfDNA analysis	cfDNA depth across genomic bins	CNVs; rare autosomal trisomies	Classification and segmentation	Screening	Low per-bin signal; maternal CNVs and malignancy	Bianchi <i>et al.</i> , 2015; Merhala <i>et al.</i> , 2026; Turriff <i>et al.</i> , 2024
Non-invasive monogenic genotyping	Deep cfDNA sequencing with parental genotypes	Inherited and <i>de novo</i> SNVs and small indels	Genotype inference (Bayesian with RF; DL)	Screening or diagnosis in at-risk families	Maternal allele background; sequencing depth; small cohorts	Liscovitch-Brauer <i>et al.</i> , 2024; Qi <i>et al.</i> , 2024; Schwammenthal <i>et al.</i> , 2025
Image-based genetic risk	Two-dimensional ultrasound (face, nuchal region); ear photographs	Aneuploidy; syndromic disorders	Classification; landmark measurement	Screening triage	Non-specific phenotype; case-control enrichment; operator dependence	Hennocq <i>et al.</i> , 2024; Tang, Han, Xue, <i>et al.</i> , 2023; L. Zhang <i>et al.</i> , 2022
Structural anomaly detection	Fetal echocardiography; anomaly scan	Indirect; phenotypes with genetic aetiology	Classification	Referral for diagnostic genetic testing	Genetic yield varies by phenotype	Arnaout <i>et al.</i> , 2021; D'Alberti <i>et al.</i> , 2025; Lord <i>et al.</i> , 2019
Automated karyotyping	Metaphase images from amniocytes or villi	Aneuploidy; structural rearrangements; mosaicism	Segmentation and classification	Laboratory diagnostic aid	Rarity of structural abnormalities and low-level mosaicism in training data	Qin <i>et al.</i> , 2019; Wu <i>et al.</i> , 2026; Zhu <i>et al.</i> , 2025
Variant calling and pathogenicity prediction	Fetal exome or genome reads	SNVs; indels; splice-altering variants	Classification and scoring	Laboratory diagnostic aid	Calibration to evidence strength; ancestry bias	Cheng <i>et al.</i> , 2023; Jaganathan <i>et al.</i> , 2019; Pejaver <i>et al.</i> , 2022; Poplin <i>et al.</i> , 2018
Prioritisation and CNV classification	Variants with HPO-coded fetal phenotype; CNV calls	SNVs; CNVs	Prioritisation; classification	Laboratory diagnostic aid	Sparse fetal phenotypes; postnatal training data	Ardiles-Ruesjas <i>et al.</i> , 2026; Meng <i>et al.</i> , 2023; Schuetz <i>et al.</i> , 2024
Language-model support	Clinical text and images	All classes (interpretive)	Generation	Decision support; drafting	Fabricated content; no prenatal validation	Suhag <i>et al.</i> , 2023; Zhong <i>et al.</i> , 2025

Note. The table is illustrative of the principal application types and is not exhaustive. AI = artificial intelligence; cfDNA = cell-free DNA; CNN = convolutional neural network; CNV = copy number variant; DL = deep learning; HPO = Human Phenotype Ontology; RF = random forest; SNV = single-nucleotide variant; SVM = support vector machine

3.2 Analytical Tasks and Clinical Roles

Across these signal sources, AI performs five broad analytical tasks. Estimation tasks quantify a continuous parameter, such as fetal fraction, on which downstream testing depends. Classification tasks assign a sample, image or chromosome to a category, such as trisomic or euploid, or normal or abnormal. Genotype inference reconstructs the fetal allele state at specific loci from mixed maternal and fetal signal. Prioritisation ranks candidate variants or diagnoses for expert review. Generation, the task performed by LLMs, produces text such as differential diagnoses, phenotype annotations or draft reports. The clinical role of these tasks also differs. An algorithm may act as a quality-control step, as a screening test that raises or lowers the probability of disease, as a diagnostic aid within a laboratory workflow or as a decision-support tool for counselling. The distinction between screening and diagnosis is fundamental in prenatal care, because a screening result, however refined, requires confirmation by diagnostic testing before irreversible decisions are made (Dungan *et al.*, 2023).

The performance metrics that are meaningful also differ by role. For screening, detection rate and false-positive rate must be interpreted alongside prevalence, because the PPV of even a highly specific test falls sharply in low-risk populations. The contrast between the PPV of cfDNA testing and of conventional screening in a routine population (Norton *et al.*, 2015) illustrates that the relevant comparator for a new AI screening tool is the best available current test in the intended population, not an obsolete or single-marker benchmark. For laboratory tasks such as chromosome classification or variant prioritisation, the relevant metrics are accuracy at the level of the case, the proportion of true diagnoses ranked within a reviewable shortlist and the time saved without loss of sensitivity (Meng *et al.*, 2023; Qin *et al.*, 2019).

Table 2 organises the principal AI applications according to signal source, variant class, analytical task, clinical role and the dominant constraint that limits performance. The table shows that the constraint most likely to limit accuracy differs by domain. It is biological in cfDNA analysis, phenotypic in imaging, related to rare-event representation in cytogenetics and interpretive in sequencing. Transferring performance expectations from one domain to another is therefore unjustified.

4. Artificial Intelligence in Cell-Free DNA Analysis

4.1 Fetal Fraction Estimation as the Foundation of Non-Invasive Inference

Fetal fraction, the proportion of cfDNA in maternal plasma that is of placental origin, governs the sensitivity of every cfDNA-based inference because the magnitude of a fetal aneuploidy or variant signal scales with it. In pregnancies with a male fetus, fetal fraction can be estimated from the proportion of reads mapping to the Y chromosome, but no equivalent direct marker exists in female pregnancies. This practical gap was the first setting in which multivariate learning became integral to NIPT. The SeqFF method trained a multivariate model on regional autosomal read counts from 25,312 pregnancies and, applied to independent samples, correlated with Y-chromosome estimates at Pearson coefficients of 0.932 to 0.938 and with a single-nucleotide polymorphism method at 0.921 (Kim *et al.*, 2015). The model exploited the non-uniform genomic distribution of shorter cfDNA fragments, which are enriched for fetal origin, illustrating how learning approaches can extract information from features that are biologically plausible but not individually decisive. A neural network method, FF-QuantSC, subsequently reported high correlation with Y-based estimates and comparable behaviour in male and female pregnancies using routine shallow sequencing data (Yuan *et al.*, 2020). Reviews of fetal fraction estimation describe a family of count-based, fragment-size-based, methylation-based and polymorphism-based approaches, each with characteristic strengths and limitations (Peng & Jiang, 2017; Tsui *et al.*, 2025).

The evidence for these methods is analytically robust in the narrow sense that they reproduce reference estimates in large datasets. Their clinical significance is less secure. A laboratory evaluation of 14,379 shallow-sequenced samples found that bioinformatic choices had a substantial effect on fetal fraction, with different tools performing best in male and female pregnancies (Hestand *et al.*, 2019). An international external quality assessment found wide variability in fetal fraction reported for identical samples across laboratories, with coefficients of variation between 17.0% and 35.8%, and a systematic difference between two widely used estimators in 87,351 clinical samples (Becking *et al.*, 2023). The authors concluded that fetal fraction should be regarded as a laboratory-specific range rather than a precise value and that universal minimum thresholds could cause unnecessary test failures. This finding has a direct implication for AI. Estimators trained to agree with one reference method inherit its biases, and the absence of a gold standard in female pregnancies means that apparent accuracy is often agreement between methods rather than proximity to truth. Confidence that learned estimators improve clinical decisions, as distinct from analytical agreement, is therefore limited.

4.2 Aneuploidy Calling Beyond Z-Score Thresholds

Conventional NIPT classifies aneuploidy by comparing the normalised representation of a chromosome with a euploid reference, typically by a z-score threshold. ML approaches have been proposed to reduce the no-call zone, false positives and false negatives that arise near this threshold. A support vector machine combining multiple z-statistics with quality-control indices was trained on samples that the z-test could classify and was then applied to samples it could not, reporting complete accuracy in internal validation and correcting four false-positive and four false-negative z-test results (J. Yang *et al.*, 2018). A CNN applied to cfDNA fragment distance rather than fragment count was developed on 2,215 samples and evaluated on 17,678 clinical samples, in which an ensemble model achieved a sensitivity of 99.07% and a PPV of 88.43% for trisomies 21, 18 and 13, compared with PPVs of approximately 37% to 41% for conventional count-based statistics at slightly lower sensitivity (Lee *et al.*, 2022).

These results indicate that learned features can recover information that simple count statistics discard, and the scale of the second study is notable. Their interpretation nonetheless requires caution for three reasons. First, the baseline performance of cfDNA screening for common trisomies is already very high, with pooled detection rates above 97% and false-positive rates near 0.04%

(Gil *et al.*, 2017), so that absolute gains are necessarily small and are concentrated in PPV, which depends strongly on prevalence and referral pattern. The large apparent improvement in PPV in a single laboratory population may not transfer to lower-prevalence settings. Second, false-positive cfDNA results are commonly caused by CPM, maternal mosaicism or maternal disease rather than by analytical noise, so that a classifier can reduce false positives only to the extent that the relevant biological states leave distinguishable features in the data (Grati *et al.*, 2020; Merhala *et al.*, 2026). Improving concordance with fetal karyotype by suppressing calls may also suppress clinically relevant placental findings. Third, the evidence consists of retrospective single-laboratory analyses. Neither study reported prospective evaluation, independent external validation in another laboratory or analysis of calibration. The weight of evidence therefore supports ML as a plausible refinement of aneuploidy calling, but not as a demonstrated improvement in clinical outcomes.

An earlier strand of work used ML on combined first-trimester markers rather than on cfDNA. Neural network models applied to maternal age, NT and serum biochemistry reported high classification performance for trisomies and sex chromosome aneuploidies (Catic *et al.*, 2018; Jamshidnezhad *et al.*, 2022). A two-stage computational approach trained on more than 72,000 euploid pregnancies proposed contingent use of cfDNA testing, projecting a reduction in the false-positive rate from 51.4% to about 1% under the assumption that cfDNA testing identified all euploid and aneuploid cases (Neocleous *et al.*, 2018). The larger of these studies demonstrates a realistic role for ML in allocating pregnancies to cfDNA or invasive testing, particularly where universal cfDNA screening is unaffordable. The smaller studies, with a few hundred to a few thousand pregnancies and no external validation, provide weak evidence, and their comparison with conventional risk algorithms rather than with cfDNA screening limits their relevance to contemporary practice (Dungan *et al.*, 2023).

4.3 Subchromosomal and Genome-Wide Findings

Extension of cfDNA analysis from common trisomies to CNVs, microdeletions and rare autosomal trisomies is one of the most contested areas of NIPT, and AI is often presented as a means of making it reliable. The underlying difficulty is that the signal for a small CNV is spread across few genomic bins and scales with fetal fraction, while maternal CNVs, CPM and maternal malignancy introduce competing signals (Merhala *et al.*, 2026). The literature retrieved for this review contained no peer-reviewed prospective evaluation of an AI classifier for fetal CNV detection from cfDNA with confirmatory diagnostic testing across a consecutive population. The earliest evidence for learned genomic-distribution features in CNV detection is indirect, arising from the claim that fetal fraction estimators such as SeqFF could improve CNV detection (Kim *et al.*, 2015). Genome-wide screening for rare autosomal trisomies has been judged to offer minimal benefit because most such findings reflect CPM with low fetal risk, with the exception of trisomy 16 (Grati *et al.*, 2020), and professional guidance continues to regard screening for rare autosomal trisomies and microdeletions as an emerging area (Dungan *et al.*, 2023).

The maternal-malignancy literature illustrates both the promise and the hazard of pattern recognition in genome-wide cfDNA data. Specific multichromosomal patterns of gains and losses were strongly associated with occult cancer in individuals with unusual cfDNA results (Turiff *et al.*, 2024), which suggests that learned pattern classifiers could help to route non-reportable results towards maternal rather than fetal investigation. The same observation shows that genome-wide signals in pregnancy are not exclusively fetal. An AI system optimised solely for fetal outcome labels could learn to ignore, or to misattribute, maternal disease. The design of learning objectives for genome-wide NIPT therefore raises ethical as well as technical questions about which findings are sought, reported and acted upon.

4.4 Non-Invasive Genotyping of Monogenic Variants

The demonstration that maternal plasma contains the entire fetal genome established the theoretical feasibility of non-invasive diagnosis of monogenic disorders (Lo *et al.*, 2010). Early whole-genome reconstructions combined parental genome sequencing, maternal haplotyping and deep plasma sequencing, predicting fetal inheritance at 2.8 million parental heterozygous sites with 98.1% accuracy but detecting *de novo* point mutations only with limited specificity, and concluded that ultradeep sequencing would be required for genome-wide *de novo* detection (Kitzman *et al.*, 2012). Haplotype counting and exome capture of plasma DNA provided complementary routes to the inherited and *de novo* fetal exome (Fan *et al.*, 2012). Targeted non-invasive sequencing for a panel of genes associated with dominant monogenic disorders, using statistical adjustment for read count and fetal fraction rather than ML, reported 20 true-positive and 127 true-negative results with no false results among pregnancies with available outcomes (J. Zhang *et al.*, 2019). This last study provides the most clinically mature evidence for monogenic cfDNA testing and serves as a useful benchmark, because it achieved its results without learned models, within a restricted gene panel and with orthogonal confirmation of positive findings.

Against this benchmark, ML has been applied to the harder problem of genome-wide inference, particularly for maternally inherited variants, whose fetal allele must be detected against an identical maternal background. A method combining fragment features, Bayesian inference and random-forest refinement trained on millions of non-pathogenic variants reported area under the receiver operating characteristic curve (AUC) values of 0.996 for paternally inherited variants, 0.81 for maternally inherited variants, 0.86 for homozygous biallelic variants and 0.95 for compound heterozygous variants in 16 families, and correctly classified fetal status in all ten families carrying a pathogenic SNV (Liscovitch-Brauer *et al.*, 2024). The same group then described what it characterised as the first DL framework for cfDNA-based genotyping, using whole-genome sequencing of maternal plasma at a target depth of 300-fold in ten families; it reported AUC values of 0.996 to 0.999 for paternally inherited variants but 0.715 to 0.878 for maternally inherited variants, detection of three deleterious mutations and genotyping as early as the seventh week of gestation (Schwammenthal *et al.*, 2025). A self-learning neural network using 56 cfDNA and parental parameters in ten fetuses with monogenic disease reported genotype accuracies of 77.74% to 89.02% against cord-blood sequencing, varying by parental genotype configuration (Qi *et al.*, 2024).

Table 3. Representative evidence for artificial intelligence in cell-free DNA-based fetal genetic analysis, with principal limitations and evidence-maturity level

Application	AI approach	Data and validation design	Principal reported performance	Principal limitations (maturity level)	Supporting references
Fetal fraction (female and male pregnancies)	Multivariate model on regional autosomal read counts	Trained on 25,312 pregnancies; tested on independent cohorts	$r = 0.932\text{--}0.938$ versus Y-chromosome estimates; $r = 0.921$ versus SNP-based method	Agreement with reference methods rather than with a true value (Level 2)	Kim <i>et al.</i> , 2015
Fetal fraction	Neural network on genomic patterns in shallow sequencing	Male and female pregnancies from routine NIPT	High correlation with Y-based estimates; similar behaviour in female pregnancies	No gold standard in female pregnancies; wide inter-laboratory variability of estimates (Level 1–2)	Becking <i>et al.</i> , 2023; Yuan <i>et al.</i> , 2020
Trisomies 21, 18, 13	SVM combining z-statistics and quality indices	Trained on z-test-classifiable samples; tested on unclassifiable samples	Complete accuracy reported; four false-positive and four false-negative z-test results corrected	Small unresolved subset; single laboratory (Level 1)	J. Yang <i>et al.</i> , 2018
Trisomies 21, 18, 13	CNN on cfDNA fragment distance (ensemble)	Developed on 2,215 samples; evaluated on 17,678 clinical samples	Sensitivity 99.07%; PPV 88.43% versus approximately 37–41% for count-based statistics	Retrospective, single laboratory; PPV prevalence-dependent; placental mosaicism unresolved (Level 2)	Grati <i>et al.</i> , 2020; Lee <i>et al.</i> , 2022
Contingent allocation to cfDNA testing	Two-stage neural network on first-trimester markers	More than 72,000 euploid and 600 aneuploid pregnancies for training; blind set of 36,328	Trisomy 21 sensitivity 97.1% and specificity 99.5% at stage 2	Assumes perfect cfDNA performance; conventional comparator (Level 2)	Neocleous <i>et al.</i> , 2018
Inherited monogenic SNVs	Fragment features with Bayesian inference and random forest	16 families with varying fetal fraction	AUC 0.996 (paternal), 0.81 (maternal); 10 of 10 pathogenic-variant families correctly classified	Small cohort; AUC dominated by benign sites (Level 1–2)	Liscovitch-Brauer <i>et al.</i> , 2024
Genome-wide fetal genotyping	Deep neural network on whole-genome cfDNA sequencing at 300-fold target depth	10 families; held-out test data	AUC 0.996–0.999 (paternal); 0.715–0.878 (maternal)	Depth impractical for screening; no external validation (Level 1)	Schwammenthal <i>et al.</i> , 2025
Fetal SNVs in monogenic disease	Self-learning neural network with 56 cfDNA and parental parameters	10 fetuses; cord-blood sequencing reference	Genotype accuracy 77.74–89.02%; 24 of 25 high-risk SNVs confirmed	Brief report; per-variant accuracy insufficient for diagnosis (Level 1)	Qi <i>et al.</i> , 2024
Fragment origin by methylation	CNN on single-molecule polymerase kinetics; methylation along long molecules	Reference DNA and maternal plasma	Methylation AUC up to 0.97; fetal versus maternal molecule AUC 0.88	Long-read platforms not in routine screening (Level 1)	Tse <i>et al.</i> , 2021; Yu <i>et al.</i> , 2021
Congenital heart defects (non-chromosomal)	Multiple AI models on genome-wide cfDNA methylation	12 cases and 26 controls	AUC of at least 0.92; five-marker model AUC 0.97	5,918 candidate features for 12 cases; no independent validation (Level 1)	Bahado-Singh <i>et al.</i> , 2023

Note. Evidence-maturity levels: Level 1, technical development with internal validation; Level 2, external, temporal or large independent validation; Level 3, prospective clinical evaluation; Level 4, evaluation of clinical outcomes or implementation. No study reached Level 3 or 4. AI = artificial intelligence; AUC = area under the receiver operating characteristic curve; cfDNA = cell-free DNA; CNN = convolutional neural network; NIPT = non-invasive prenatal testing; PPV = positive predictive value; SNP = single-nucleotide polymorphism; SNV = single-nucleotide variant; SVM = support vector machine

These studies make a genuine methodological contribution by treating cfDNA fragment characteristics as informative features for allele origin, and their consistency in showing that maternally inherited variants remain the weakest category is itself informative. The evidence is, however, preliminary. Cohorts comprised ten to sixteen families, the plasma sequencing depth used in the DL study was far greater than the shallow coverage typical of routine aneuploidy screening (Hestand *et al.*, 2019), performance was reported as AUC across millions of mostly benign sites rather than as sensitivity and PPV for the rare pathogenic variants that matter clinically, and none of the studies reported external validation in an independent laboratory. The per-variant accuracy reported in the self-learning study, below 90% in some configurations, would be inadequate for diagnostic use without confirmatory testing. Confidence that DL can deliver genome-wide non-invasive diagnosis of monogenic disease is therefore low, although confidence that learned features improve allele-origin inference in research settings is moderate.

4.5 Epigenomic and Fragmentomic Features

A further strand uses DNA methylation and fragmentation patterns, rather than sequence alone, to infer tissue of origin and fetal status. A CNN trained on polymerase kinetic signals from single-molecule real-time sequencing detected cytosine methylation at single-base resolution with an AUC of up to 0.97 and correlated closely with bisulfite sequencing (Tse *et al.*, 2021). Applied to maternal plasma, single-molecule sequencing showed that a substantial proportion of cfDNA consists of long molecules and that methylation patterns along a long molecule could distinguish fetal from maternal origin with an AUC of 0.88, enabling inference of maternal inheritance and recombination (Yu *et al.*, 2021). These studies are mechanistically important because they provide a route to determining the parental origin of individual fragments, the central obstacle for maternally inherited variants. Their relevance to routine fetal variant detection remains prospective, since the sequencing platforms involved are not yet used for population screening.

Methylation-based AI has also been applied to non-chromosomal fetal phenotypes. A study of genome-wide cfDNA methylation in 12 pregnancies with isolated congenital heart defects and 26 controls identified 5,918 differentially methylated cytosines and reported AI models with AUCs of 0.92 or higher, including a five-marker model with an AUC of 0.97 (Bahado-Singh *et al.*, 2023). The ratio of candidate features to cases in this design creates a high risk of overfitting, and the absence of an independent validation set means that the reported discrimination is best interpreted as hypothesis-generating. The study illustrates a broader pattern in which the combination of high-dimensional omic data and small case numbers produces performance estimates that are unlikely to be reproduced.

Table 3 summarises representative evidence for AI in cfDNA-based fetal genetic analysis. The table is not exhaustive, but it shows a consistent inverse relationship between the clinical ambition of the task and the maturity of its evidence. Fetal fraction estimation and aneuploidy calling have been evaluated in tens of thousands of samples, whereas genome-wide monogenic genotyping and epigenomic prediction rest on cohorts of fewer than 40 pregnancies. The table also shows that no study reached the level of prospective evaluation with clinical outcomes.

5. Imaging-Derived Phenotypes as Proxies for Fetal Genetic Disease

5.1 From Serum-Marker Networks to Image-Based Screening

The use of AI to infer fetal genetic risk from indirect markers predates cfDNA testing. The earliest identified neural network system for this purpose combined maternal age, gestational age and second-trimester serum markers in 112 women referred for invasive testing and reported statistically significant discrimination of genetic disorders, while its authors cautioned that the system could not exceed the information supplied to it (Beksaç *et al.*, 1995). That caveat remains the central principle for image-based AI. An algorithm that analyses the fetal face, neck or heart can detect only those consequences of a genetic disorder that are expressed and visible at the time of imaging. The contemporary literature can be grouped into three strands: automated measurement of established first-trimester markers, end-to-end prediction of genetic disease from sonographic images and detection of structural anomalies that prompt genetic testing.

5.2 First-Trimester Markers and End-to-End Aneuploidy Prediction

Automated measurement of established markers is the most conservative application, because the algorithm reproduces a clinical measurement whose relationship with aneuploidy is already characterised. DL systems have been developed to identify intracranial and facial structures in mid-sagittal views at 11 to 14 weeks, reaching an AUC of 0.974 for classification of standard versus non-standard planes on external data (Q. Lin *et al.*, 2022), to assess image quality for NT measurement with an AUC of 0.978 (L. Liu *et al.*, 2025) and to measure the frontomaxillary facial angle with a mean absolute error of 2.354 degrees (Z. Liu *et al.*, 2025). A model measuring five facial profile markers achieved intraclass correlation coefficients of 0.76 to 0.81 against manual measurement in 500 normal fetuses. When applied to 18 fetuses with trisomy 21 and 7 with trisomy 18, individual markers yielded AUCs ranging from 0.686 for trisomy 21 to 0.904 for trisomy 18 (Ji *et al.*, 2023). These studies demonstrate that automation can standardise measurements that are operator-dependent. They do not show that automated markers improve aneuploidy detection over contemporary screening, and the very small numbers of affected fetuses preclude reliable estimates of marker performance.

End-to-end models attempt a more ambitious task, predicting chromosomal status directly from images without specifying markers. The most widely cited example trained a CNN on mid-sagittal images from 822 participants at two hospitals and reported an AUC of 0.95 in a validation set for trisomy 21, compared with 0.73 for a model combining NT and maternal age (L. Zhang *et al.*, 2022). The comparison appears striking, but several features of the design temper its interpretation. Cases and controls were enrolled by availability over more than a decade rather than consecutively, which enriches the sample for trisomy 21 far beyond population

prevalence. The validation set was formed by random selection from the same institutional pool rather than from an independent centre or period, so that equipment, operators and case-mix effects could not be separated from biological signal. The comparator, NT and maternal age alone, is weaker than contemporary combined screening and substantially weaker than cfDNA testing, which achieves an AUC of 0.999 for trisomy 21 in a routine population (Norton *et al.*, 2015). The claim that such a model could facilitate universal primary screening is therefore not supported by the evidence presented. A systematic review of AI in first-trimester imaging reached a consistent conclusion across 27 studies, noting that 16 used fewer than 1,000 cases and only one was externally validated (Umans *et al.*, 2024).

5.3 Facial and Morphological Gestalt in the Fetus

Postnatal facial analysis provides the conceptual template for prenatal gestalt approaches. A deep facial image analysis framework trained on more than 17,000 images covering more than 200 syndromes achieved 91% top-10 accuracy in identifying the correct syndrome from 502 images (Gurovich *et al.*, 2019). A subsequent encoder defined a clinical face phenotype space from 17,560 patients with 1,115 disorders and allowed matching of patients with ultra-rare disorders not included in training (Hsieh *et al.*, 2022). Integration of facial similarity scores with exome data improved the ranking of disease-causing genes and was proposed as a quantitative implementation of phenotypic specificity within variant classification criteria (Hsieh *et al.*, 2019). These postnatal systems work with high-resolution photographs of expressed phenotypes and large, internationally contributed datasets, conditions that do not hold for fetal ultrasound.

Prenatal adaptations have been reported mainly from a single research programme. A DL model trained on fetal facial ultrasound images from 1,000 pregnancies, covering trisomies 21, 18 and 13, Turner syndrome, several microdeletion and duplication syndromes and monogenic disorders, reported sensitivities of 0.75 to 0.96 and specificities of 0.92 to 0.95 across categories, with performance comparable to experienced sonographers (Tang, Han, Xue, *et al.*, 2023). The published report indicates that the data came from a single centre, that augmentation expanded 970 training images to 3,340, that 186 original images were used for testing and that no external test set was available, a limitation acknowledged by the authors. A three-stage model from the same programme, trained on 2,554 images and using detection of facial regions followed by ensemble classification, reported a sensitivity of 0.753 and specificity of 0.889 internally, and a sensitivity of 0.768 and specificity of 0.837 on an external test set (Tang, Han, Jiang, *et al.*, 2023). The lower external specificity is consistent with the expected effect of dataset shift and suggests that the internal estimates from the first model are optimistic. An automated ear-shape model trained on 1,489 photographs from 526 children and tested on photographs of 51 fetuses examined in fetal pathology distinguished CHARGE syndrome, mandibulofacial dysostosis of the Guion-Almeida type and controls with an overall accuracy of 72.6%, and the authors cautioned that further validation was required before diagnostic use (Hennocq *et al.*, 2024). This study is valuable as a transparent example of postnatal-to-fetal transfer, but it concerns fetal pathology rather than in-utero detection.

Several sources of disagreement between prenatal gestalt studies can be identified. Genetic labels combine conditions with very different phenotypic expression, so that aggregate sensitivity depends on the mix of trisomies, CNVs and monogenic disorders included. Most studies are case-control designs in which affected fetuses were ascertained because they had abnormal imaging or invasive testing, which inflates the prevalence of visible phenotypes among cases. Augmentation of small datasets can improve robustness but does not add independent biological information, and random image-level splits can place images from the same fetus in both training and test sets unless the partition is performed at the level of the pregnancy. Facial morphology also varies by ancestry and gestational age, and none of the fetal gestalt studies retrieved for this review reported performance across ancestral groups. Confidence in current claims that DL can detect rare genetic disorders from fetal facial ultrasound is therefore low.

5.4 Structural Anomaly Detection as a Gateway to Genetic Testing

A more indirect but more mature route by which imaging AI contributes to the detection of fetal genetic disease is the identification of structural anomalies that trigger diagnostic genetic testing. Congenital heart disease (CHD) is the paradigm. An ensemble of neural networks trained on 107,823 images from 1,326 studies distinguished normal hearts from complex CHD with an AUC of 0.99 and a sensitivity of 95% in an internal test set of 4,108 fetal surveys, and remained robust on outside-hospital images (Arnaout *et al.*, 2021). In a subsequent evaluation using community-acquired images from a low-risk Dutch population, in which clinical screening had detected only 53% of isolated severe CHD, the model achieved a sensitivity of 91% and a specificity of 78%, whereas blinded experts reviewing the same stored images achieved means of 55% and 71% (Athalye *et al.*, 2024). A meta-analysis of 15 studies estimated pooled sensitivity and specificity of 0.89 and 0.91 for binary normal-versus-abnormal classification, while rating the risk of bias as moderate to high and noting that human comparator groups were small and usually not blinded (D'Alberti *et al.*, 2025). Reviews of AI in obstetric ultrasound and fetal magnetic resonance imaging describe similar developments for the central nervous system and other organ systems, largely at the development stage (Horgan *et al.*, 2023; Meshaka *et al.*, 2023).

The relevance of these findings to genetic detection lies in the downstream yield. In a large prospective cohort, diagnostic genetic variants were identified by exome sequencing in 11.1% of fetuses with cardiac anomalies and 15.4% of those with multisystem anomalies (Lord *et al.*, 2019). Improved anomaly detection therefore increases the number of fetuses in whom a genetic diagnosis is possible, but it does not itself detect variants, and the benefit depends on the availability of timely diagnostic testing. The case-control composition of the external CHD evaluation, with 66 affected and 42 unaffected fetuses, means that its specificity estimate cannot be translated directly into a population false-positive rate. A registered international multicentre protocol has explicitly identified the retrospective nature of most studies and the scarcity of diverse, well-curated data as barriers to clinical use (Patey *et al.*, 2025).

Table 4. Representative imaging-based artificial intelligence studies relevant to fetal genetic detection, contrasted by target, design, validation and principal limitations

Target and task	Imaging input	Data and validation design	Principal reported performance	Principal limitations (maturity level)	Supporting references
Genetic disorders from serum markers (historical)	None (maternal age, gestational age, triple test)	112 referred women	Statistically significant discrimination	Small referred sample; obsolete markers (Level 1)	Beksaç <i>et al.</i> , 1995
Standard plane and intracranial structures, 11–14 weeks	Mid-sagittal two-dimensional ultrasound	1,528 images; external test at a second site	External AUC 0.974 for standard-plane classification	Image-quality task rather than genetic prediction (Level 2)	Q. Lin <i>et al.</i> , 2022
NT image quality and frontomaxillary facial angle	Mid-sagittal ultrasound	2,140 and 1,549 images, respectively	AUC 0.978 for quality assessment; mean absolute angle error 2.354 degrees	Measurement or quality-control support only (Level 1)	L. Liu <i>et al.</i> , 2025; Z. Liu <i>et al.</i> , 2025
Facial profile markers for trisomies 21 and 18	Mid-sagittal profile images	2,372 normal fetuses; 18 with trisomy 21 and 7 with trisomy 18	ICC 0.76–0.81; AUC 0.686 (trisomy 21) to 0.904 (trisomy 18) for single markers	Very few affected fetuses (Level 1)	Ji <i>et al.</i> , 2023
End-to-end trisomy 21 prediction	Mid-sagittal face and nuchal region	822 cases and controls from two hospitals; random split	Validation AUC 0.95 versus 0.73 for NT with maternal age	Enriched case-control design; weak comparator; no independent site (Level 1)	L. Zhang <i>et al.</i> , 2022
Facial gestalt for 12 genetic disorders	Fetal facial ultrasound	1,000 pregnancies; augmentation of 970 to 3,340 training images; 186 test images; single centre	Sensitivity 0.75–0.96; specificity 0.92–0.95	No external test set (Level 1)	Tang, Han, Xue, <i>et al.</i> , 2023
Facial region ensemble for genetic disorders	Fetal facial ultrasound	2,554 images; internal and external test sets	External sensitivity 0.768 and specificity 0.837	Specificity lower on external data (Level 2)	Tang, Han, Jiang, <i>et al.</i> , 2023
CHARGE and mandibulofacial dysostosis	External ear photographs	Trained on 1,489 photographs of 526 children; tested on 51 fetuses in fetal pathology	Overall accuracy 72.6%	Fetal pathology rather than in-utero imaging; small test set (Level 2)	Hennocq <i>et al.</i> , 2024
Complex CHD as a gateway phenotype	Fetal screening ultrasound and echocardiography	107,823 images from 1,326 studies; internal test of 4,108 surveys	AUC 0.99; sensitivity 95%; specificity 96%	Tertiary-centre training data (Level 2)	Arnaout <i>et al.</i> , 2021
Isolated severe CHD in a low-risk population	Stored community-acquired images	66 CHD and 42 normal fetuses	Sensitivity 91% and specificity 78% versus 53% clinical detection	Case-control composition; stored rather than live images (Level 2)	Athalye <i>et al.</i> , 2024
Fetal cardiac abnormalities (meta-analysis)	Screening ultrasound and echocardiography	15 studies; 30,121 fetuses	Pooled sensitivity 0.89 and specificity 0.91	Moderate-to-high risk of bias; low reporting adherence	D'Alberti <i>et al.</i> , 2025

Note. Maturity levels as defined in Table 3. Performance values are those reported by the original authors and are not directly comparable across rows because targets, prevalence and reference standards differ. AUC = area under the receiver operating characteristic curve; CHD = congenital heart disease; ICC = intraclass correlation coefficient; NT = nuchal translucency

Table 4 contrasts representative imaging-based AI studies according to their target, design, validation and principal limitations. The comparison highlights a divergence between the two imaging strands. Studies of structural anomaly detection have begun to include external and community-acquired data and have been synthesised by meta-analysis. Studies predicting chromosomal or syndromic status directly from images remain predominantly single-centre, enriched case-control designs with internal validation only. Both strands share a reliance on stored images, which differ from the dynamic information available during a live scan.

6. Artificial Intelligence in Prenatal Cytogenetics

6.1 Automated Chromosome Segmentation and Classification

G-banded karyotyping remains part of prenatal diagnosis because it provides whole-genome visualisation of banding patterns at single-cell resolution, yet it is labour-intensive and depends on specialised expertise for the manual analysis of metaphase spreads (Zhu *et al.*, 2025). AI has been applied to the most time-consuming steps, namely segmentation of overlapping chromosomes in metaphase spreads and classification of each chromosome by type and orientation. A two-scale CNN approach evaluated on 1,909 karyotyping cases reported case-level accuracy of 99.2% for chromosome type and polarity and was stated to have been applied in practical karyotype diagnosis (Qin *et al.*, 2019). The release of a clinical dataset comprising 126,453 G-banded chromosome instances from 2,763 karyotypes enabled benchmarking across methods and a baseline accuracy of 99.33% (C. Lin *et al.*, 2022), and an open dataset of annotated metaphase images supported a chromosome detector with 98.88% identification accuracy (Tseng *et al.*, 2023). Methods that reduce the need for manually drawn masks and that address batch effects and domain shift have also been developed and deployed clinically (Chen *et al.*, 2024).

Prenatal-specific evidence has emerged more recently. A modular workflow trained on 13,223 clinical amniotic fluid cases, comprising more than 50,000 manually annotated chromosomes, maintained classification accuracy of 97.45%, 96.95% and 95.72% across training, temporal validation and independent testing, and generated draft karyotypes from unsorted metaphase images with close concordance to expert review (Wu *et al.*, 2026). A cytogenetic laboratory evaluation of AI-assisted karyotyping on peripheral blood described a two-stage implementation framework in which algorithmic output is reviewed by cytogeneticists before reporting (Zhu *et al.*, 2025). A review of commercial AI-guided karyotyping products described their entry into clinical laboratories and the resulting changes to workflow and staff roles (Rosenblum *et al.*, 2025).

6.2 From Chromosome Accuracy to Detection of Abnormal Karyotypes

The high accuracies reported in this domain require careful interpretation because the analytical unit is usually the individual normal chromosome, not the abnormal fetus. A classifier that correctly assigns 99% of chromosomes may still fail on the rare derivative, marker or ring chromosome that determines the diagnosis. Direct evaluation of structural abnormalities is limited. A single-stage DL model trained on 2,424 normal and 544 abnormal chromosomes reported 91.75% accuracy for classification of the 24 normal chromosome types, 87.76% for a 32-class task including abnormal forms and 90.84% to 100% for eight common structural abnormalities (C. Yang *et al.*, 2023). The lower accuracy of the multiclass task, compared with benchmarks of normal chromosome classification, illustrates the difficulty of learning rare abnormal morphologies from limited examples. The prenatal amniotic fluid study reported accuracy at the chromosome level and concordance of draft karyotypes, but the published abstract does not report sensitivity for abnormal karyotypes, mosaicism or balanced rearrangements separately (Wu *et al.*, 2026).

Mosaicism poses a particular challenge. Low-level mosaicism in cultured amniocytes is detected by examining sufficient cells to exclude a minority abnormal line, and AI-assisted metaphase selection could, in principle, either increase efficiency or bias cell selection towards morphologically typical spreads. No peer-reviewed English-language study retrieved for this review reported sensitivity of AI-assisted prenatal karyotyping for low-level mosaicism against an orthogonal reference standard. The prevailing model of deployment, in which AI proposes a karyotype that is verified by an expert (Rosenblum *et al.*, 2025; Zhu *et al.*, 2025), mitigates the risk of missed abnormalities but also means that reported gains are gains in efficiency rather than in diagnostic sensitivity. The evidence supports AI-assisted karyotyping as a validated efficiency tool under human oversight, with moderate confidence, but does not establish that it can independently detect abnormal fetal karyotypes. The relative role of karyotyping is also changing, as chromosomal microarray and sequencing assume primary positions for many prenatal indications (Riggs *et al.*, 2020; Van den Veyver *et al.*, 2022), which limits the long-term clinical impact of gains in this domain.

7. Artificial Intelligence in the Interpretation of Fetal Sequencing Data

7.1 Variant Calling and Computational Pathogenicity Prediction

Once fetal DNA has been sequenced, AI contributes at two levels: detecting variants in the raw data and estimating their functional consequences. A deep CNN that learns from images of read pileups outperformed established variant callers for SNVs and small insertions and deletions and generalised across sequencing technologies and experimental designs (Poplin *et al.*, 2018). A deep neural network trained on pre-messenger RNA sequence predicted splice junctions from primary sequence, enabling identification of cryptic splice-altering variants, and its authors estimated that 9% to 11% of pathogenic mutations in rare genetic disorders act through this mechanism (Jaganathan *et al.*, 2019). AlphaMissense, adapted from a protein-structure model and fine-tuned on population frequency data, provided pathogenicity predictions for all possible human single amino acid substitutions and classified 89% of missense variants as likely benign or likely pathogenic (Cheng *et al.*, 2023).

These tools are not prenatal-specific, but they operate within every fetal exome and genome pipeline. Their incorporation into clinical classification is governed by the framework of the American College of Medical Genetics and Genomics and the

Association for Molecular Pathology, in which computational evidence contributes at a supporting level unless calibrated otherwise (Richards *et al.*, 2015). Calibration of computational predictors against independent data has shown that most tools can provide supporting evidence, while several reach moderate or strong evidence levels within defined score intervals (Pejaver *et al.*, 2022). This calibration step is critical for the prenatal context. A high predictor score is not a diagnosis, and in a setting where findings may inform decisions about pregnancy continuation, overstating the weight of computational evidence carries particular risk. A major source of uncertainty in fetal sequencing is the interpretation of variants with suggestive but insufficient evidence. In the two prospective cohorts described earlier, such variants were identified in a further 3.9% and 20% of fetuses, respectively, alongside diagnostic yields of 8.5% and 10% (Lord *et al.*, 2019; Petrovski *et al.*, 2019). Computational predictors reduce but do not eliminate this uncertainty.

7.2 Phenotype-Driven Prioritisation Under Sparse Fetal Phenotypes

Variant prioritisation integrates genotype with phenotype, typically encoded in the Human Phenotype Ontology (HPO), to rank candidate variants for expert review. The ontology has expanded continuously and supports semantic similarity and ML algorithms across clinical diagnostics (Gargano *et al.*, 2024), and efforts to include fetal phenotypes have been described as essential for prenatal application (Shear *et al.*, 2025). In postnatal rare disease, automated prioritisation has reached a level of maturity that supports routine laboratory use. An ML model trained on thousands of manually curated variants ranked the reported diagnostic variant within the top ten candidates in 96.7% of 180 previously solved exome cases and was implemented in a prospective production cohort, although accuracy declined when CNV calling was incomplete or phenotypic descriptions were limited (Meng *et al.*, 2023). A reanalysis strategy based on a phenotype-driven prioritisation tool achieved a recall of 82% and a precision of 88% for new candidates in the United Kingdom 100,000 Genomes Project, with automated classification correctly upgrading 92% of relevant variants (Vestito *et al.*, 2024).

Transfer of these results to the fetus is problematic. Fetal phenotypes are recorded by ultrasound or magnetic resonance imaging at a single point in development, many features such as intellectual disability or dysmorphism are inaccessible, and phenotypes evolve with gestation (Shear *et al.*, 2025). A large prenatal ES cohort illustrates the consequences. In 1,618 fetuses with structural anomalies and normal karyotype and microarray results, a genotype-driven first step identified 11.6% of diagnoses, a phenotype-driven second step added 1.7% and reanalysis based on new phenotypic information added a further 0.9% (Fu *et al.*, 2022). The study also expanded the known prenatal phenotype for dozens of genes, confirming that fetal presentations are incompletely represented in knowledge bases used to train and tune prioritisation algorithms. A retrospective comparison in 206 pregnancies found that HPO-driven gene selection would have detected only 71% of causative genes identified by interpreting morbid genes broadly, compared with 96% for a curated virtual panel (Ardiles-Ruesjas *et al.*, 2026). This result directly challenges the assumption that phenotype-driven filtering, the principle underlying most AI prioritisation, is well suited to fetal cases.

The apparent conflict between postnatal and prenatal performance is explicable. Postnatal algorithms are trained and evaluated on cases with rich, mature phenotypes, whereas fetal phenotypes are sparse, non-specific and sometimes misleading. Algorithms that weight phenotype similarity heavily may deprioritise the correct gene when the fetal presentation is atypical or incompletely described. Aggregation of standardised prenatal genotype and phenotype data has been proposed as a prerequisite for tuning algorithms to the prenatal space (Duyzend *et al.*, 2024). Image-based phenotyping could, in principle, complement HPO terms in the way that facial analysis complements postnatal exome interpretation (Hsieh *et al.*, 2019), but prenatal equivalents remain unvalidated. No peer-reviewed study retrieved for this review reported the performance of an automated prioritisation system on a prospectively ascertained prenatal cohort with ranking of diagnostic variants as the primary endpoint, and registry searches identified no registered evaluation of AI for prenatal sequencing interpretation.

7.3 Copy Number Variant Classification and Reporting Automation

CNV interpretation follows a points-based framework that integrates genomic content, dosage sensitivity, overlap with established syndromes and inheritance (Riggs *et al.*, 2020). ML has been used to predict clinical significance directly from CNV features. A framework trained on 52,176 ClinVar entries and validated against 21,574 CNVs from an external database outperformed alternative approaches and provided explanations for its predictions (Schuetz *et al.*, 2024). Its training labels, however, derive from ClinVar submissions that are not specific to prenatal presentations, and its performance on prenatal CNVs, including those of uncertain significance and variable penetrance, was not evaluated. In a prenatal and pregnancy-loss laboratory, natural language processing was applied to 30,235 historical reports to build knowledge bases for automated CNV reporting, achieving variant classification accuracy of 95.7% in retrospective, 95.2% in prospective and 100% in clinical utility experiments (M. Liu *et al.*, 2020). This study is among the few to report prospective evaluation of an AI tool in a prenatal genetic laboratory, although its knowledge base is inherently local and would require re-derivation in other laboratories.

7.4 Large Language Models in Prenatal Genetic Interpretation

LLMs have attracted interest as aids to differential diagnosis, phenotype annotation and patient communication. A commentary on the use of a general-purpose chatbot in complex prenatal differential diagnosis highlighted both its potential for generating diagnostic hypotheses from de-identified findings and its limitations (Suhag *et al.*, 2023). The most directly relevant empirical evidence concerns phenotype encoding. In a comparative study of multimodal LLMs, junior physicians assisted by model-suggested HPO terms achieved 67.4% accuracy compared with 20.4% for manual searching, yet the models alone generated incorrect ontology identifiers in more than half of their outputs and fabricated terms in about one-third (Zhong *et al.*, 2025). Because phenotype terms feed directly into variant prioritisation, such errors could propagate silently through the interpretive pipeline. A commentary on clinical acumen in prenatal diagnosis concluded that AI is promising as an adjunct but does not replace the clinical

skills needed for detailed imaging, integration of history and informed differential diagnosis (Sparks & Chitty, 2026). At present, the evidence supports LLMs only as supervised aids for drafting and retrieval, with low confidence in any diagnostic role in prenatal genetics.

8. Cross-Cutting Appraisal: Validity, Generalisability, Equity and Governance

8.1 Reference Standards, Verification and the Meaning of a Label

The validity of any diagnostic AI study depends on the reference standard used to define its labels. In prenatal genetics, the reference standard is rarely uniform. Fetal karyotype or microarray results are usually available only for pregnancies that screened positive or had anomalies, while screen-negative pregnancies are classified by newborn examination or remain unverified. This pattern of differential verification is familiar from conventional screening studies, in which birth outcome was established by diagnostic testing or newborn examination (Norton *et al.*, 2015), and it is amplified in AI studies that draw retrospectively on clinical databases. Algorithms trained on such data may learn the referral pathway as well as the disease. In cfDNA studies, the question of whether the label should be fetal karyotype or placental karyotype is unresolved, because CPM produces true placental abnormalities that are false with respect to the fetus (Grati *et al.*, 2020). In imaging studies, genetic labels are available only for fetuses that underwent invasive testing, which enriches cases for visible anomalies. In sequencing studies, the reference is the laboratory's own classification, which changes with reanalysis (Fu *et al.*, 2022; Vestito *et al.*, 2024). The reporting guideline for AI-centred diagnostic accuracy studies emphasises description of dataset practices, the reference standard and algorithmic bias (Sounderajah *et al.*, 2025), but few of the prenatal studies reviewed here would satisfy those items in full.

8.2 Dataset Shift, External Validation and Reporting Quality

Across domains, the most consistent weakness is the absence of independent external validation. A systematic review of first-trimester imaging AI identified only one externally validated model among 27 (Umans *et al.*, 2024), and a meta-analysis of cardiac imaging AI reported low adherence to prediction-model reporting items and a moderate-to-high risk of bias (D'Alberti *et al.*, 2025). The prediction-model reporting guideline for regression and ML methods requires explicit description of data sources, handling of missing data, sample size, performance measures including calibration, and fairness considerations (Collins *et al.*, 2024). Guidance on early clinical evaluation emphasises human factors, the interaction between users and algorithmic outputs and safety monitoring (Vasey *et al.*, 2022), and guidance for randomised trials of AI interventions specifies reporting of the algorithm version, input data quality and handling of errors (X. Liu *et al.*, 2020). Calibration was not reported in the published abstracts of most primary prenatal studies examined here, and none reported a randomised or prospective comparative evaluation with patient-level outcomes.

The consequences of limited validation differ by domain. In cfDNA analysis, differences in library preparation, sequencing platforms and bioinformatic pipelines shift the distributions of read counts and fragment features, and the inter-laboratory variability in fetal fraction estimates (Becking *et al.*, 2023) demonstrates that such shifts are substantial. In imaging, differences in equipment, operators and scanning protocols alter image characteristics, and the decline in specificity observed on external data in fetal facial analysis (Tang, Han, Jiang, *et al.*, 2023) is consistent with this effect. In cytogenetics, variation in banding quality and metaphase preparation has motivated methods explicitly designed to handle domain shift (Chen *et al.*, 2024). The broader literature on clinical AI has identified these issues, together with performance metrics that do not reflect clinical applicability, as central barriers to deployment (Kelly *et al.*, 2019).

8.3 Ancestry, Equity and Access

Genomic reference resources remain dominated by populations of European ancestry (Popejoy & Fullerton, 2016), and the practical harm of this imbalance has been documented. In hypertrophic cardiomyopathy testing, patients of African or unspecified ancestry received positive reports based on variants that were later reclassified as benign, and the misclassified variants were significantly more common among Black Americans than among White Americans (Manrai *et al.*, 2016). AI tools for variant interpretation inherit the composition of the data used to train and calibrate them, and prenatal decisions are especially sensitive to false-positive pathogenic classifications. For image-based models, facial and craniofacial morphology varies by ancestry, and the fetal gestalt studies reviewed here were each derived from a single national population without reported subgroup analysis. For cfDNA algorithms, maternal body mass index and other biological factors influence fetal fraction (Hestand *et al.*, 2019), so that performance can vary systematically across groups that differ in their distributions.

Equity also concerns access. The most advanced AI applications, including deep non-invasive genotyping and single-molecule methylation analysis, depend on sequencing resources that are unavailable in most health systems. Conversely, AI support for ultrasound has been proposed as a way to extend expertise to non-specialist or low-resource settings (Patey *et al.*, 2025), where detection rates for fetal anomalies are lowest. Whether AI narrows or widens disparities in prenatal genetic diagnosis will depend on where it is validated and deployed. Guidance from the World Health Organization identifies inclusiveness and equity among six core principles for AI in health (World Health Organization, 2021).

8.4 Ethical, Counselling and Accountability Considerations

Prenatal genetic information is used in decisions that are time-limited, emotionally charged and sometimes irreversible. An ethical analysis of genomic AI in prenatal and paediatric medicine identified concerns relating to beneficence, non-maleficence, autonomy, justice, transparency, accountability, privacy and trust, and noted that these applications had received little attention in the ethics literature (Coghlan *et al.*, 2024). Three issues are particularly salient to the evidence reviewed here. The first is opacity. When a DL

model assigns a raised probability of a genetic syndrome from a facial image or a cfDNA profile, clinicians and parents may be unable to understand the basis of the output, which complicates counselling and informed consent. Heatmaps and feature-attribution plots offer partial explanations (Schuetz *et al.*, 2024; Tang, Han, Xue, *et al.*, 2023) but do not establish that the model relies on biologically valid features.

The second issue is the scope of findings. Algorithms that extend cfDNA or sequencing analysis genome-wide will generate incidental and secondary findings, including adult-onset conditions and maternal disease (Fu *et al.*, 2022; Turriff *et al.*, 2024). The position statement of the International Society for Prenatal Diagnosis emphasises standardised laboratory practice and detailed pre-test and post-test counselling for prenatal sequencing (Van den Veyver *et al.*, 2022), and these requirements apply with greater force when automated systems widen the range of detectable findings. The third issue is accountability. Current deployments place AI outputs under expert review (Rosenblum *et al.*, 2025; Zhu *et al.*, 2025), which preserves professional responsibility but may encourage automation bias if reviewers come to accept algorithmic proposals uncritically. The fabrication of ontology terms by LLMs (Zhong *et al.*, 2025) exemplifies the type of error that is difficult for reviewers to detect when outputs appear plausible.

Table 5 synthesises the recurring methodological limitations identified across domains, their likely consequences for reported performance and the design features that could address them. The table shows that most limitations lead to overestimation of performance rather than underestimation, which supports a conservative reading of current accuracy claims.

Table 5. Recurring methodological limitations in studies of artificial intelligence for fetal genetic detection, their likely consequences and design remedies

Limitation	Domains most affected	Likely consequence for reported performance	Design remedy	Supporting references
Case-control enrichment and non-consecutive sampling	Image-based aneuploidy and syndrome prediction; CHD evaluation; epigenomic cfDNA	Inflated predictive values not transferable to screening populations	Consecutive cohorts at intended-use prevalence	Athalye <i>et al.</i> , 2024; Bahado-Singh <i>et al.</i> , 2023; L. Zhang <i>et al.</i> , 2022
Internal validation only; partitioning at image rather than pregnancy level	First-trimester imaging; fetal facial gestalt; monogenic cfDNA	Optimistic discrimination; possible leakage between training and test data	Pregnancy-level partitioning; temporal and external validation	Schwammenthal <i>et al.</i> , 2025; Tang, Han, Xue, <i>et al.</i> , 2023; Umans <i>et al.</i> , 2024
Placental and maternal signal not modelled	cfDNA aneuploidy and CNV analysis	Errors attributed to algorithms that reflect biology; suppression of relevant findings	Separate fetal, placental and maternal outcome labels	Grati <i>et al.</i> , 2020; Merhala <i>et al.</i> , 2026; Turriff <i>et al.</i> , 2024
Agreement with a reference method treated as accuracy	Fetal fraction estimation	Transfer of reference-method bias; overconfidence	Multi-method comparison; laboratory-specific ranges	Becking <i>et al.</i> , 2023; Kim <i>et al.</i> , 2015
Chromosome-level rather than case-level metrics	Automated karyotyping	High accuracy masking missed rare abnormalities and mosaicism	Case-level sensitivity for abnormal karyotypes and mosaicism	Wu <i>et al.</i> , 2026; C. Yang <i>et al.</i> , 2023
Postnatal training data and sparse fetal phenotypes	Variant prioritisation; CNV classification	Lower ranking performance and missed diagnoses in fetuses	Prenatal benchmark cohorts; standardised fetal phenotyping	Ardiles-Ruesjas <i>et al.</i> , 2026; Duyzend <i>et al.</i> , 2024; Meng <i>et al.</i> , 2023
Weak or outdated comparators	Imaging and marker-based aneuploidy models	Exaggerated incremental value	Comparison with cfDNA screening and current combined screening	Neocleous <i>et al.</i> , 2018; Norton <i>et al.</i> , 2015
Ancestry imbalance and absent subgroup reporting	Variant interpretation; facial models	Differential misclassification and inequitable error rates	Diverse datasets; stratified performance reporting	Manrai <i>et al.</i> , 2016; Popejoy & Fullerton, 2016
Incomplete reporting and absent calibration	All domains	Limited reproducibility and uncertain applicability	Adherence to TRIPOD+AI, STARD-AI and DECIDE-AI	Collins <i>et al.</i> , 2024; D'Alberty <i>et al.</i> , 2025; Sounderajah <i>et al.</i> , 2025
Fabricated or unverifiable generated content	Language-model phenotype annotation	Silent propagation of erroneous terms into prioritisation	Constrained outputs; mandatory human verification	Zhong <i>et al.</i> , 2025

Note. CHD = congenital heart disease; cfDNA = cell-free DNA; CNV = copy number variant; DECIDE-AI = reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence; STARD-AI = Standards for Reporting Diagnostic Accuracy for artificial intelligence; TRIPOD+AI = Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis, updated for artificial intelligence

9. Integrative Synthesis: Where Artificial Intelligence Adds Value and Where Claims Outrun Evidence

Taken together, the evidence supports a differentiated rather than uniform judgement. The strongest evidence concerns tasks in which AI operates as an analytical component within an established testing pathway, a measurable reference exists and expert oversight is retained. Fetal fraction estimation in female pregnancies (Kim *et al.*, 2015; Yuan *et al.*, 2020), chromosome classification in cytogenetic laboratories (Qin *et al.*, 2019; Wu *et al.*, 2026) and variant prioritisation in rare-disease genomics (Meng *et al.*, 2023; Vestito *et al.*, 2024) fall into this category. The benefits in these areas are chiefly efficiency, consistency and scalability, and they are achieved without the algorithm making an unsupervised determination of fetal genetic status.

An intermediate category comprises tasks in which AI refines an existing screening decision. ML-based aneuploidy calling from cfDNA (Lee *et al.*, 2022; J. Yang *et al.*, 2018) and AI-supported detection of structural anomalies that trigger genetic testing (Arnaout *et al.*, 2021; Athalye *et al.*, 2024) have shown credible analytical gains in large retrospective or external datasets. Their clinical benefit remains unproven because prospective evaluations with appropriate comparators are lacking and because the dominant sources of error, placental mosaicism and phenotypic variability, are biological rather than computational.

The weakest evidence concerns tasks in which AI is proposed as a new primary detector of fetal genetic disease. End-to-end prediction of aneuploidy or syndromes from ultrasound images (Tang, Han, Xue, *et al.*, 2023; L. Zhang *et al.*, 2022), genome-wide non-invasive genotyping of monogenic variants (Liscovitch-Brauer *et al.*, 2024; Qi *et al.*, 2024; Schwammmenthal *et al.*, 2025) and epigenomic prediction of fetal anomalies from cfDNA (Bahado-Singh *et al.*, 2023) rest on small, enriched or single-centre datasets with internal validation. These studies are scientifically valuable as proofs of concept, and several introduce novel signal representations, but their reported performance should not be used to justify clinical adoption.

Fig. 1 presents a conceptual framework, developed from the present synthesis rather than adapted from prior work, that relates each signal source to the AI tasks applied to it, the variant classes addressed and the evidence maturity reached. Solid connections denote relationships supported by external, temporal or large independent validation, and dashed connections denote relationships supported only by internal validation or proof-of-concept studies. The framework makes explicit two structural features of the field. First, the closer an AI task lies to the final diagnostic decision, the thinner its evidence base. Second, each signal source carries a biological ceiling, placental origin for cfDNA, phenotypic expression for imaging, cell sampling for cytogenetics and knowledge-base completeness for sequencing, that no algorithm can exceed. Improvements in detection will therefore require better signals and better reference standards as much as better models.

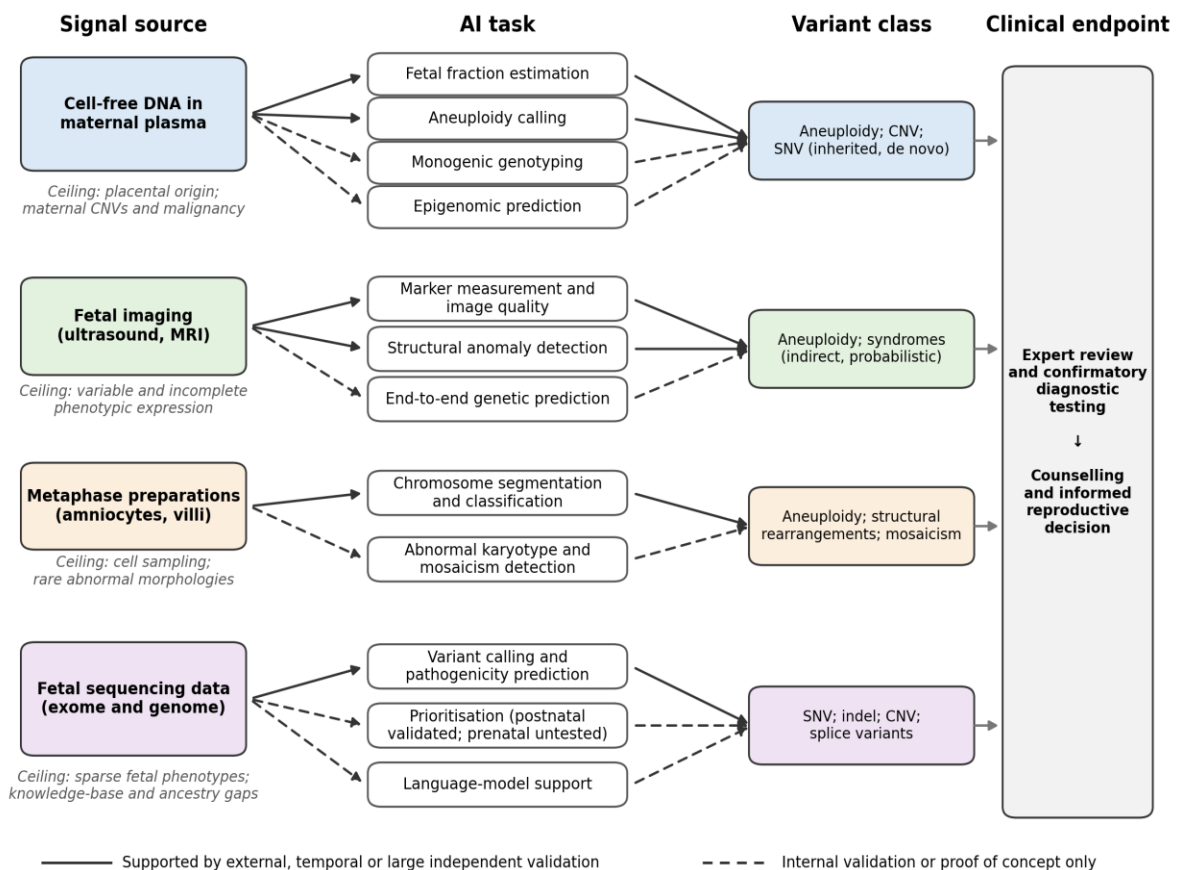


Fig. 1. Conceptual framework relating the signal sources used for prenatal genetic detection to the artificial intelligence (AI) tasks applied to them, the variant classes addressed and the maturity of supporting evidence. The framework was developed from the present synthesis and is not adapted from prior work. Solid arrows denote relationships supported by external, temporal or large independent validation; dashed arrows denote relationships supported only by internal validation or proof-of-concept studies. The italic text beneath each signal source states its principal biological or interpretive ceiling. All outputs converge on expert review and confirmatory diagnostic testing before counselling. The relationships shown are hypothesised to hold across settings but have not been tested prospectively. CNV = copy number variant; MRI = magnetic resonance imaging; SNV = single-nucleotide variant

10. Future Directions

The most immediate priority is prospective, consecutive-cohort evaluation of AI tools that are already technically mature. For ML-based aneuploidy calling and fetal fraction estimation, the unresolved problem is whether analytical gains translate into fewer invasive procedures, fewer failed tests and preserved sensitivity in routine populations. Current evidence is inadequate because it is retrospective, derived from single laboratories and reported as discrimination rather than as clinical outcomes (Becking *et al.*, 2023; Lee *et al.*, 2022). Multicentre studies in unselected screening populations, with pre-specified comparison against the laboratory's current pipeline, confirmation of screen-positive results by diagnostic testing and follow-up of screen-negative pregnancies to birth, would address this gap. Relevant outcome measures include PPV at intended-use prevalence, test-failure rate, invasive procedure rate, time to result and calibration across gestational age and maternal body mass index. Because cfDNA pipelines differ between laboratories, external validation should be performed on data generated with different library preparation and sequencing protocols. Such studies would allow professional guidance to specify whether and under what conditions learned models can replace conventional statistical calling.

A second immediate priority concerns reference standards for cfDNA-based AI. Algorithms should be evaluated against fetal, placental and maternal outcomes separately wherever possible, so that learned features can be attributed to their biological source. Paired analysis of placental and fetal samples in pregnancies with discordant results, together with systematic follow-up of maternal findings, would establish whether AI can distinguish CPM, maternal CNVs and maternal malignancy from true fetal abnormality (Grati *et al.*, 2020; Turriff *et al.*, 2024). This work is essential before genome-wide AI-supported NIPT for CNVs is offered as a screening test, because the ethical acceptability of such screening depends on knowing what an abnormal result means.

For non-invasive monogenic testing, the principal research need is scale. Existing ML and DL studies include ten to sixteen families (Liscovitch-Brauer *et al.*, 2024; Schwammenthal *et al.*, 2025), and their performance for maternally inherited variants remains insufficient for diagnosis. Multicentre consortia should assemble family-based datasets with invasive or postnatal confirmation, report sensitivity and PPV for pathogenic variants rather than aggregate AUC across benign sites and evaluate performance at sequencing depths that are economically feasible. Integration of long-read and methylation-based origin inference (Tse *et al.*, 2021; Yu *et al.*, 2021) is a longer-term direction that could resolve the maternal-background problem, but it requires cost reduction before population relevance can be assessed.

For imaging-based genetic risk prediction, the immediate need is methodological discipline rather than new architectures. Studies should use consecutive screening cohorts at realistic prevalence, partition data at the level of the pregnancy, validate externally across centres, equipment and ancestral groups and compare performance against contemporary cfDNA-based screening rather than against single markers (Norton *et al.*, 2015; Umans *et al.*, 2024). The most plausible role for such models is as triage tools that identify fetuses whose phenotype warrants diagnostic testing even after a low-risk cfDNA result, a role that should be tested explicitly. For structural anomaly detection, randomised or stepped-wedge implementation studies in non-specialist settings, such as those foreshadowed by current multicentre protocols (Patey *et al.*, 2025), should measure prenatal detection rates, referral appropriateness and downstream genetic diagnostic yield.

For sequencing interpretation, the priority is a prenatal benchmark. Standardised, shared datasets of fetal exomes and genomes with structured phenotypes and confirmed diagnoses (Duyzend *et al.*, 2024) would allow prioritisation algorithms, CNV classifiers and *in silico* predictors to be evaluated on the population in which they are used. Outcome measures should include the rank of the diagnostic variant, the proportion of diagnoses missed by phenotype-driven filtering (Ardiles-Ruesjas *et al.*, 2026), turnaround time and the rate of variants of uncertain significance reported. Calibration of computational predictors to evidence strength within the prenatal context, and assessment of performance across ancestries (Manrai *et al.*, 2016; Pejaver *et al.*, 2022), should be part of this programme. LLM applications should be restricted to supervised tasks and evaluated with explicit measurement of fabrication and error propagation (Zhong *et al.*, 2025).

Longer-term priorities concern governance and human factors. Studies following early clinical evaluation guidance (Vasey *et al.*, 2022) should examine how clinicians, laboratory scientists and genetic counsellors interact with AI outputs, including automation bias and the communication of algorithmic uncertainty to parents. Qualitative and mixed-methods research with pregnant individuals and families should examine how AI-derived risk estimates influence understanding, anxiety and decision-making. Regulatory and professional frameworks should require post-deployment monitoring of performance drift and equity, consistent with international guidance on AI for health (World Health Organization, 2021). Progress on these fronts would shift the field from demonstrating that algorithms can classify prenatal data to demonstrating that they improve prenatal genetic care.

11. Conclusions

This review set out to define the tasks performed by artificial intelligence across the prenatal genetic testing pathway, to appraise the strength of the supporting evidence and to identify the sources of disagreement and uncertainty. The synthesis indicates that artificial intelligence is already embedded in fetal genetic detection, but mainly as an analytical component rather than as an autonomous detector. Its best-supported contributions are the estimation of fetal fraction, the classification of chromosomes in cytogenetic workflows and the prioritisation of candidate variants for expert review. In these roles, algorithms operate against measurable references and under professional oversight, and their benefits are principally those of efficiency and consistency.

Machine-learning refinement of aneuploidy calls from cell-free DNA and deep-learning detection of structural anomalies that prompt genetic testing have produced credible analytical gains in large retrospective and external datasets. Their clinical value remains uncertain because prospective evaluations in unselected populations have not been reported and because the principal

sources of error in these domains are biological. Placental mosaicism and maternal confounders limit what any algorithm can learn from plasma DNA, and the variable and incomplete expression of genetic disease in the fetus limits what any algorithm can learn from images.

Claims that artificial intelligence can detect fetal aneuploidy or rare syndromes directly from ultrasound images, or reconstruct pathogenic fetal genotypes genome-wide from maternal plasma, rest on small, enriched or single-centre studies with internal validation. These studies are valuable as demonstrations of feasibility but do not justify clinical use. The interpretation of fetal sequencing data is constrained less by algorithmic capacity than by sparse fetal phenotypes, postnatal training data and ancestry imbalance in reference resources, and large language models introduce additional risks of fabricated content.

The broader significance of this synthesis lies in the observation that the limiting factors in prenatal genetic detection are the nature of the signal, the adequacy of the reference standard and the representativeness of the data, rather than the architecture of the model. Artificial intelligence is therefore most defensibly positioned as an adjunct that strengthens established testing pathways, subject to transparent reporting, external validation, equitable datasets and governance that protects informed reproductive choice.

12. Limitations

As a narrative review, this article is subject to selection and interpretive bias despite a transparent search strategy and explicit appraisal criteria. Studies were selected for relevance, methodological adequacy and representativeness rather than enumerated exhaustively, and the thematic organisation reflects judgements about which distinctions matter most. Other reviewers could reasonably weight the evidence differently, particularly where individual studies straddle categories of evidence maturity.

Searching relied on openly accessible bibliographic databases, a trial registry and citation tracing. Subscription-only indexes were not searched, so relevant publications indexed exclusively in those sources may have been missed, particularly engineering and conference literature. Publications that were not in English were not included unless their bibliographic records could be fully verified, and relevant work published in other languages, including a substantial body of research from East Asia, may therefore be under-represented. Appraisal of some primary studies depended on published abstracts and accessible full texts, which did not always report the information needed to judge sampling, data partitioning or calibration.

The evidence base is heterogeneous in its data modalities, analytical units, reference standards and outcome measures, which precluded quantitative synthesis and limits the comparability of performance figures across studies. Publication bias is likely, because algorithm development studies with favourable results are more readily published than negative or null evaluations, and this would lead to overestimation of performance. Geographical representation is uneven, with many primary studies originating from a small number of health systems. The field is developing rapidly, and new models, datasets and evaluations published after the search end date may alter some of the conclusions drawn here. Long-term evidence on clinical outcomes, family experience and equity effects of prenatal genetic AI is essentially absent, so conclusions about benefit and harm remain provisional.

Declaration of AI Use

This manuscript was prepared through the intellectual contributions of the author(s) to the study design, data analysis, interpretation of results, and scholarly content. Grammarly and ChatGPT were used solely to assist with grammatical refinement and reference formatting during manuscript revision. The author(s) accept full responsibility for the accuracy, integrity, and final content of the manuscript.

Competing Interests

Authors have declared that no competing interests exist.

References

- Ardiles-Ruesjas, V., Rodriguez-Reventa, L., Pauta, M., Nadal, A., Arca, G., Gort, L., Sinisterra-Díaz, S. E., & Borrell, A. (2026). Gene list selection matters: Missed diagnoses in prenatal exome sequencing—PanelApp R21 and HPO-driven versus OMIM-based gene lists. *Prenatal Diagnosis*, 46(5–6), 636–642. <https://doi.org/10.1002/pd.70066>
- Arnaout, R., Curran, L., Zhao, Y., Levine, J. C., Chinn, E., & Moon-Grady, A. J. (2021). An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. *Nature Medicine*, 27(5), 882–891. <https://doi.org/10.1038/s41591-021-01342-5>
- Athalye, C., van Nesselrooij, A., Rizvi, S., Haak, M. C., Moon-Grady, A. J., & Arnaout, R. (2024). Deep-learning model for prenatal congenital heart disease screening generalizes to community setting and outperforms clinical detection. *Ultrasound in Obstetrics & Gynecology*, 63(1), 44–52. <https://doi.org/10.1002/uog.27503>
- Baethge, C., Goldbeck-Wood, S., & Mertens, S. (2019). SANRA—a scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review*, 4(1), Article 5. <https://doi.org/10.1186/s41073-019-0064-8>
- Bahado-Singh, R., Friedman, P., Talbot, C., Aydas, B., Southekal, S., Mishra, N. K., Guda, C., Yilmaz, A., Radhakrishna, U., & Vishweswaraiah, S. (2023). Cell-free DNA in maternal blood and artificial intelligence: Accurate prenatal detection of fetal congenital heart defects. *American Journal of Obstetrics and Gynecology*, 228(1), 76.e1–76.e10. <https://doi.org/10.1016/j.ajog.2022.07.062>

- Basel-Salmon, L., & Brabbing-Goldstein, D. (2024). Fetal whole genome sequencing as a clinical diagnostic tool: Advantages, limitations and pitfalls. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 97, Article 102549. <https://doi.org/10.1016/j.bpobgyn.2024.102549>
- Becking, E. C., Linthorst, J., Patton, S., Gutowska-Ding, W., Goodall, R., Khawaja, F., Morgan, F., Deans, Z., Chitty, L. S., Bekker, M. N., Scheffer, P. G., & Sistermans, E. A. (2023). Variability in fetal fraction estimation: Comparing fetal fractions reported by noninvasive prenatal testing providers globally. *Clinical Chemistry*, 69(2), 160–167. <https://doi.org/10.1093/clinchem/hvac207>
- Beksaç, M. S., Durak, B., Özkan, Ö., Çakar, A. N., Balci, S., Karakaş, Ü., & Laleli, Y. (1995). An artificial intelligent diagnostic system with neural networks to determine genetical disorders and fetal health by using maternal serum markers. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 59(2), 131–136. [https://doi.org/10.1016/0028-2243\(94\)02034-C](https://doi.org/10.1016/0028-2243(94)02034-C)
- Bianchi, D. W., Chudova, D., Sehner, A. J., Bhatt, S., Murray, K., Prosen, T. L., Garber, J. E., Wilkins-Haug, L., Vora, N. L., Warsof, S., Goldberg, J., Ziainia, T., & Halks-Miller, M. (2015). Noninvasive prenatal testing and incidental detection of occult maternal malignancies. *JAMA*, 314(2), 162–169. <https://doi.org/10.1001/jama.2015.7120>
- Boddupally, K., & Rani Thuraka, E. (2024). Artificial intelligence for prenatal chromosome analysis. *Clinica Chimica Acta*, 552, Article 117669. <https://doi.org/10.1016/j.cca.2023.117669>
- Bouachba, A., De Jesus Neves, J., Royer, E., Martin, R., Salomon, L. J., Grevent, D., & Gorincour, G. (2025). Artificial intelligence, radiomics and fetal ultrasound: Review of literature and future perspectives. *Ultrasound in Obstetrics & Gynecology*, 65(3), 281–291. <https://doi.org/10.1002/uog.29172>
- Catic, A., Gurbeta, L., Kurtovic-Kozaric, A., Mehmedbasic, S., & Badnjevic, A. (2018). Application of neural networks for classification of Patau, Edwards, Down, Turner and Klinefelter syndrome based on first trimester maternal serum screening data, ultrasonographic findings and patient demographics. *BMC Medical Genomics*, 11(1), Article 19. <https://doi.org/10.1186/s12920-018-0333-2>
- Chen, S., Zhang, K., Hu, J., Li, N., Xu, A., Li, H., Zhou, J., Huang, C., Yu, Y., & Gao, X. (2024). KaryoXpert: An accurate chromosome segmentation and classification framework for karyotyping analysis without training with manually labeled metaphase-image mask annotations. *Computers in Biology and Medicine*, 177, Article 108601. <https://doi.org/10.1016/j.combiomed.2024.108601>
- Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L. H., Zielinski, M., Sargeant, T., Schneider, R. G., Senior, A. W., Jumper, J., Hassabis, D., Kohli, P., & Avsec, Ž. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*, 381(6664), Article eadg7492. <https://doi.org/10.1126/science.adg7492>
- Chiu, R. W. K., Chan, K. C. A., Gao, Y., Lau, V. Y. M., Zheng, W., Leung, T. Y., Foo, C. H. F., Xie, B., Tsui, N. B. Y., Lun, F. M. F., Zee, B. C. Y., Lau, T. K., Cantor, C. R., & Lo, Y. M. D. (2008). Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proceedings of the National Academy of Sciences of the United States of America*, 105(51), 20458–20463. <https://doi.org/10.1073/pnas.0810641105>
- Coghlan, S., Gyngell, C., & Vears, D. F. (2024). Ethics of artificial intelligence in prenatal and pediatric genomic medicine. *Journal of Community Genetics*, 15(1), 13–24. <https://doi.org/10.1007/s12687-023-00678-4>
- Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A.-L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., . . . Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385, Article e078378. <https://doi.org/10.1136/bmj-2023-078378>
- D'Alberty, E., Patey, O., Smith, C., Šalović, B., Hernandez-Cruz, N., Noble, J. A., & Papageorgiou, A. T. (2025). Artificial intelligence-enabled prenatal ultrasound for the detection of fetal cardiac abnormalities: A systematic review and meta-analysis. *eClinicalMedicine*, 84, Article 103250. <https://doi.org/10.1016/j.eclinm.2025.103250>
- Dungan, J. S., Klugman, S., Darilek, S., Malinowski, J., Akkari, Y. M. N., Monaghan, K. G., Erwin, A., Best, R. G., & ACMG Board of Directors. (2023). Noninvasive prenatal screening (NIPS) for fetal chromosome abnormalities in a general-risk population: An evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 25(2), Article 100336. <https://doi.org/10.1016/j.gim.2022.11.004>
- Duong, D., & Solomon, B. D. (2025). Artificial intelligence in clinical genetics. *European Journal of Human Genetics*, 33(3), 281–288. <https://doi.org/10.1038/s41431-024-01782-w>
- Duyzend, M. H., Cacheiro, P., Jacobsen, J. O. B., Giordano, J., Brand, H., Wapner, R. J., Talkowski, M. E., Robinson, P. N., & Smedley, D. (2024). Improving prenatal diagnosis through standards and aggregation. *Prenatal Diagnosis*, 44(4), 454–464. <https://doi.org/10.1002/pd.6522>
- Fan, H. C., Gu, W., Wang, J., Blumenfeld, Y. J., El-Sayed, Y. Y., & Quake, S. R. (2012). Non-invasive prenatal measurement of the fetal genome. *Nature*, 487(7407), 320–324. <https://doi.org/10.1038/nature11251>
- Fu, F., Li, R., Yu, Q., Wang, D., Deng, Q., Li, L., Lei, T., Chen, G., Nie, Z., Yang, X., Han, J., Pan, M., Zhen, L., Zhang, Y., Jing, X., Li, F., Li, F., Zhang, L., Yi, C., . . . Liao, C. (2022). Application of exome sequencing for prenatal diagnosis of fetal structural anomalies: Clinical experience and lessons learned from a cohort of 1618 fetuses. *Genome Medicine*, 14(1), Article 123. <https://doi.org/10.1186/s13073-022-01130-x>
- Gargano, M. A., Matentzoglou, N., Coleman, B., Addo-Lartey, E. B., Anagnostopoulos, A. V., Anderton, J., Avillach, P., Bagley, A. M., Bakstein, E., Balhoff, J. P., Baynam, G., Bello, S. M., Berk, M., Bertram, H., Bishop, S., Blau, H., Bodenstein, D. F., Botas, P., Boztug, K., . . . Robinson, P. N. (2024). The Human Phenotype Ontology in 2024: Phenotypes around the world. *Nucleic Acids Research*, 52(D1), D1333–D1346. <https://doi.org/10.1093/nar/gkad1005>
- Gil, M. M., Accurti, V., Santacruz, B., Plana, M. N., & Nicolaides, K. H. (2017). Analysis of cell-free DNA in maternal blood in screening for aneuploidies: Updated meta-analysis. *Ultrasound in Obstetrics & Gynecology*, 50(3), 302–314. <https://doi.org/10.1002/uog.17484>

- Grati, F. R., Ferreira, J., Benn, P., Izzi, C., Verdi, F., Vercellotti, E., Dalpiaz, C., D'Ajello, P., Filippi, E., Volpe, N., Malvestiti, F., Maggi, F., Simoni, G., Frusca, T., Cirelli, G., Bracalente, G., Lo Re, A., Surico, D., Ghi, T., & Prefumo, F. (2020). Outcomes in pregnancies with a confined placental mosaicism and implications for prenatal screening using cell-free DNA. *Genetics in Medicine*, 22(2), 309–316. <https://doi.org/10.1038/s41436-019-0630-y>
- Greenhalgh, T., Thorne, S., & Malterud, K. (2018). Time to challenge the spurious hierarchy of systematic over narrative reviews? *European Journal of Clinical Investigation*, 48(6), Article e12931. <https://doi.org/10.1111/eci.12931>
- Gurovich, Y., Hanani, Y., Bar, O., Nadav, G., Fleischer, N., Gelbman, D., Basel-Salmon, L., Krawitz, P. M., Kamphausen, S. B., Zenker, M., Bird, L. M., & Gripp, K. W. (2019). Identifying facial phenotypes of genetic disorders using deep learning. *Nature Medicine*, 25(1), 60–64. <https://doi.org/10.1038/s41591-018-0279-0>
- Hennocq, Q., Garcelon, N., Bongibault, T., Bouygues, T., Marlin, S., Amiel, J., Boutaud, L., Douillet, M., Lyonnet, S., Pingault, V., Picard, A., Rio, M., Attie-Bitach, T., Khonsari, R. H., & Roux, N. (2024). Artificial intelligence-based diagnosis in fetal pathology using external ear shapes. *Prenatal Diagnosis*, 44(10), 1150–1158. <https://doi.org/10.1002/pd.6577>
- Hestand, M. S., Bessem, M., van Rijn, P., de Menezes, R. X., Sie, D., Bakker, I., Boon, E. M. J., Sistermans, E. A., & Weiss, M. M. (2019). Fetal fraction evaluation in non-invasive prenatal screening (NIPS). *European Journal of Human Genetics*, 27(2), 198–202. <https://doi.org/10.1038/s41431-018-0271-7>
- Horgan, R., Nehme, L., & Abuhamad, A. (2023). Artificial intelligence in obstetric ultrasound: A scoping review. *Prenatal Diagnosis*, 43(9), 1176–1219. <https://doi.org/10.1002/pd.6411>
- Hsieh, T.-C., Bar-Haim, A., Moosa, S., Ehmke, N., Gripp, K. W., Pantel, J. T., Danyel, M., Mensah, M. A., Horn, D., Rosnev, S., Fleischer, N., Bonini, G., Hustinx, A., Schmid, A., Knaus, A., Javanmardi, B., Klinkhammer, H., Lesmann, H., Sivalingam, S., . . . Krawitz, P. M. (2022). GestaltMatcher facilitates rare disease matching using facial phenotype descriptors. *Nature Genetics*, 54(3), 349–357. <https://doi.org/10.1038/s41588-021-01010-x>
- Hsieh, T.-C., Mensah, M. A., Pantel, J. T., Aguilar, D., Bar, O., Bayat, A., Becerra-Solano, L., Bentzen, H. B., Biskup, S., Borisov, O., Braaten, O., Ciaccio, C., Coutelier, M., Cremer, K., Danyel, M., Daschkey, S., Eden, H. D., Devriendt, K., Wilson, S., . . . Krawitz, P. M. (2019). PEDIA: Prioritization of exome data by image analysis. *Genetics in Medicine*, 21(12), 2807–2814. <https://doi.org/10.1038/s41436-019-0566-2>
- Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki, J. A., Arbelaez, J., Cui, W., Schwartz, G. B., Chow, E. D., Kanterakis, E., Gao, H., Kia, A., Batzoglu, S., Sanders, S. J., & Farh, K. K.-H. (2019). Predicting splicing from primary sequence with deep learning. *Cell*, 176(3), 535–548.e24. <https://doi.org/10.1016/j.cell.2018.12.015>
- Jamshidnezhad, A., Hosseini, S. M., Mahmudi, M., & Mohammadi-Asl, J. (2022). A machine learning technology to improve the risk of non-invasive prenatal tests. *Technology and Health Care*, 30(4), 951–965. <https://doi.org/10.3233/THC-213628>
- Ji, C., Liu, K., Yang, X., Cao, Y., Cao, X., Pan, Q., Yang, Z., Sun, L., Yin, L., Deng, X., & Ni, D. (2023). A novel artificial intelligence model for fetal facial profile marker measurement during the first trimester. *BMC Pregnancy and Childbirth*, 23(1), Article 718. <https://doi.org/10.1186/s12884-023-06046-x>
- Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17(1), Article 195. <https://doi.org/10.1186/s12916-019-1426-2>
- Kim, S. K., Hannum, G., Geis, J., Tynan, J., Hogg, G., Zhao, C., Jensen, T. J., Mazloom, A. R., Oeth, P., Ehrich, M., van den Boom, D., & Decu, C. (2015). Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. *Prenatal Diagnosis*, 35(8), 810–815. <https://doi.org/10.1002/pd.4615>
- Kitzman, J. O., Snyder, M. W., Ventura, M., Lewis, A. P., Qiu, R., Simmons, L. E., Gammill, H. S., Rubens, C. E., Santillan, D. A., Murray, J. C., Tabor, H. K., Bamshad, M. J., Eichler, E. E., & Shendure, J. (2012). Noninvasive whole-genome sequencing of a human fetus. *Science Translational Medicine*, 4(137), Article 137ra76. <https://doi.org/10.1126/scitranslmed.3004323>
- Lee, J., Lee, S.-M., Ahn, J. M., Lee, T.-R., Kim, W., Cho, E.-H., & Ki, C.-S. (2022). Development and performance evaluation of an artificial intelligence algorithm using cell-free DNA fragment distance for non-invasive prenatal testing (aiD-NIPT). *Frontiers in Genetics*, 13, Article 999587. <https://doi.org/10.3389/fgene.2022.999587>
- Lin, C., Chen, H., Huang, J., Peng, J., Guo, L., Yang, Z., Du, J., Li, S., Yin, A., & Zhao, G. (2022). ChromosomeNet: A massive dataset enabling benchmarking and building baselines of clinical chromosome classification. *Computational Biology and Chemistry*, 100, Article 107731. <https://doi.org/10.1016/j.compbiolchem.2022.107731>
- Lin, Q., Zhou, Y., Shi, S., Zhang, Y., Yin, S., Liu, X., Peng, Q., Huang, S., Jiang, Y., Cui, C., She, R., Xu, J., & Dong, F. (2022). How much can AI see in early pregnancy: A multi-center study of fetus head characterization in week 10–14 in ultrasound using deep learning. *Computer Methods and Programs in Biomedicine*, 226, Article 107170. <https://doi.org/10.1016/j.cmpb.2022.107170>
- Liscovitch-Brauer, N., Mesika, R., Rabinowitz, T., Volkov, H., Grad, M., Matar, R. T., Basel-Salmon, L., Tadmor, O., Beker, A., & Shomron, N. (2024). Machine learning-enhanced noninvasive prenatal testing of monogenic disorders. *Prenatal Diagnosis*, 44(9), 1024–1032. <https://doi.org/10.1002/pd.6570>
- Liu, L., Wang, T., Zhu, W., Zhang, H., Tian, H., Li, Y., Cai, W., & Yang, P. (2025). Intelligent quality assessment of ultrasound images for fetal nuchal translucency measurement during the first trimester of pregnancy based on deep learning models. *BMC Pregnancy and Childbirth*, 25(1), Article 741. <https://doi.org/10.1186/s12884-025-07863-y>
- Liu, M., Zhong, Y., Liu, H., Liang, D., Liu, E., Zhang, Y., Tian, F., Liang, Q., Cram, D. S., Wang, H., Wu, L., & Yu, F. (2020). REDBot: Natural language process methods for clinical copy number variation reporting in prenatal and products of conception diagnosis. *Molecular Genetics & Genomic Medicine*, 8(11), Article e1488. <https://doi.org/10.1002/mgg3.1488>
- Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J., Denniston, A. K., & SPIRIT-AI and CONSORT-AI Working Group. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *Nature Medicine*, 26(9), 1364–1374. <https://doi.org/10.1038/s41591-020-1034-x>
- Liu, Z., Wang, J., Lyu, G., Song, H., Yu, W., Liu, P., Fan, Y., & Wan, Y. (2025). Automatic measurement of frontomaxillary facial angle in fetal ultrasound images using deep learning. *Sensors*, 25(3), Article 633. <https://doi.org/10.3390/s25030633>

- Lo, Y. M. D., Chan, K. C. A., Sun, H., Chen, E. Z., Jiang, P., Lun, F. M. F., Zheng, Y. W., Leung, T. Y., Lau, T. K., Cantor, C. R., & Chiu, R. W. K. (2010). Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Science Translational Medicine*, 2(61), Article 61ra91. <https://doi.org/10.1126/scitranslmed.3001720>
- Lo, Y. M. D., Corbetta, N., Chamberlain, P. F., Rai, V., Sargent, I. L., Redman, C. W. G., & Wainscoat, J. S. (1997). Presence of fetal DNA in maternal plasma and serum. *The Lancet*, 350(9076), 485–487. [https://doi.org/10.1016/S0140-6736\(97\)02174-0](https://doi.org/10.1016/S0140-6736(97)02174-0)
- Lord, J., McMullan, D. J., Eberhardt, R. Y., Rinck, G., Hamilton, S. J., Quinlan-Jones, E., Prigmore, E., Keelagher, R., Best, S. K., Carey, G. K., Mellis, R., Robart, S., Berry, I. R., Chandler, K. E., Cilliers, D., Cresswell, L., Edwards, S. L., Gardiner, C., Henderson, A., . . . Prenatal Assessment of Genomes and Exomes Consortium. (2019). Prenatal exome sequencing analysis in fetal structural anomalies detected by ultrasonography (PAGE): A cohort study. *The Lancet*, 393(10173), 747–757. [https://doi.org/10.1016/S0140-6736\(18\)31940-8](https://doi.org/10.1016/S0140-6736(18)31940-8)
- Manrai, A. K., Funke, B. H., Rehm, H. L., Olesen, M. S., Maron, B. A., Szolovits, P., Margulies, D. M., Loscalzo, J., & Kohane, I. S. (2016). Genetic misdiagnoses and the potential for health disparities. *The New England Journal of Medicine*, 375(7), 655–665. <https://doi.org/10.1056/NEJMs1507092>
- Mellis, R., Oprych, K., Scotchman, E., Hill, M., & Chitty, L. S. (2022). Diagnostic yield of exome sequencing for prenatal diagnosis of fetal structural anomalies: A systematic review and meta-analysis. *Prenatal Diagnosis*, 42(6), 662–685. <https://doi.org/10.1002/pd.6115>
- Meng, L., Attali, R., Talmy, T., Regev, Y., Mizrahi, N., Smirin-Yosef, P., Vossaert, L., Taborda, C., Santana, M., Machol, I., Xiao, R., Dai, H., Eng, C., Xia, F., & Tzur, S. (2023). Evaluation of an automated genome interpretation model for rare disease routinely used in a clinical genetic laboratory. *Genetics in Medicine*, 25(6), Article 100830. <https://doi.org/10.1016/j.gim.2023.100830>
- Merhala, D., Veszprémi, B., & Vass, R. A. (2026). Copy number variant detection by NIPT: Biological constraints and the limits of prenatal genomic inference. *Genes*, 17(6), Article 636. <https://doi.org/10.3390/genes17060636>
- Meshaka, R., Gaunt, T., & Shelmerdine, S. C. (2023). Artificial intelligence applied to fetal MRI: A scoping review of current research. *The British Journal of Radiology*, 96(1147), Article 20211205. <https://doi.org/10.1259/bjr.20211205>
- Neocleous, A. C., Syngelaki, A., Nicolaides, K. H., & Schizas, C. N. (2018). Two-stage approach for risk estimation of fetal trisomy 21 and other aneuploidies using computational intelligence systems. *Ultrasound in Obstetrics & Gynecology*, 51(4), 503–508. <https://doi.org/10.1002/uog.17558>
- Norton, M. E., Jacobsson, B., Swamy, G. K., Laurent, L. C., Ranzini, A. C., Brar, H., Tomlinson, M. W., Pereira, L., Spitz, J. L., Holleman, D., Cuckle, H., Musci, T. J., & Wapner, R. J. (2015). Cell-free DNA analysis for noninvasive examination of trisomy. *The New England Journal of Medicine*, 372(17), 1589–1597. <https://doi.org/10.1056/NEJMoa1407349>
- Patey, O., Hernandez-Cruz, N., D'Alberty, E., Salovic, B., Noble, J. A., Papageorgiou, A. T., & CAIFE Research Group. (2025). Prenatal detection of congenital heart defects using the deep learning-based image and video analysis: Protocol for Clinical Artificial Intelligence in Fetal Echocardiography (CAIFE), an international multicentre multidisciplinary study. *BMJ Open*, 15(6), Article e101263. <https://doi.org/10.1136/bmjopen-2025-101263>
- Pejaver, V., Byrne, A. B., Feng, B.-J., Pagel, K. A., Mooney, S. D., Karchin, R., O'Donnell-Luria, A., Harrison, S. M., Tavtigian, S. V., Greenblatt, M. S., Biesecker, L. G., Radivojac, P., Brenner, S. E., & ClinGen Sequence Variant Interpretation Working Group. (2022). Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *The American Journal of Human Genetics*, 109(12), 2163–2177. <https://doi.org/10.1016/j.ajhg.2022.10.013>
- Peng, X. L., & Jiang, P. (2017). Bioinformatics approaches for fetal DNA fraction estimation in noninvasive prenatal testing. *International Journal of Molecular Sciences*, 18(2), Article 453. <https://doi.org/10.3390/ijms18020453>
- Petrovski, S., Aggarwal, V., Giordano, J. L., Stosic, M., Wou, K., Bier, L., Spiegel, E., Brennan, K., Stong, N., Jobanputra, V., Ren, Z., Zhu, X., Mebane, C., Nahum, O., Wang, Q., Kamalakaran, S., Malone, C., Anyane-Yeboah, K., Miller, R., . . . Wapner, R. J. (2019). Whole-exome sequencing in the evaluation of fetal structural anomalies: A prospective cohort study. *The Lancet*, 393(10173), 758–767. [https://doi.org/10.1016/S0140-6736\(18\)32042-7](https://doi.org/10.1016/S0140-6736(18)32042-7)
- Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature*, 538(7624), 161–164. <https://doi.org/10.1038/538161a>
- Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., & DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36(10), 983–987. <https://doi.org/10.1038/nbt.4235>
- Qi, Y., Hu, C., Yang, J., Gao, Y., & Yin, A. (2024). Self-learning neural network as a prediction model in non-invasive prenatal testing to detect fetal SNVs. *Journal of Translational Medicine*, 22(1), Article 707. <https://doi.org/10.1186/s12967-024-05433-y>
- Qin, Y., Wen, J., Zheng, H., Huang, X., Yang, J., Song, N., Zhu, Y.-M., Wu, L., & Yang, G.-Z. (2019). Varifocal-Net: A chromosome classification approach using deep convolutional networks. *IEEE Transactions on Medical Imaging*, 38(11), 2569–2581. <https://doi.org/10.1109/TMI.2019.2905841>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., & ACMG Laboratory Quality Assurance Committee. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424. <https://doi.org/10.1038/gim.2015.30>
- Riggs, E. R., Andersen, E. F., Cherry, A. M., Kantarci, S., Kearney, H., Patel, A., Raca, G., Ritter, D. I., South, S. T., Thorland, E. C., Pineda-Alvarez, D., Aradhya, S., & Martin, C. L. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genetics in Medicine*, 22(2), 245–257. <https://doi.org/10.1038/s41436-019-0686-8>

- Rosenblum, L. S., Holmes, J., & Taghiyev, A. F. (2025). The emergence of artificial intelligence-guided karyotyping: A review and reflection. *Genes*, 16(6), Article 685. <https://doi.org/10.3390/genes16060685>
- Schuetz, R. J., Ceyhan, D., Antoniou, A. A., Chaudhari, B. P., & White, P. (2024). CNVoyant a machine learning framework for accurate and explainable copy number variant classification. *Scientific Reports*, 14(1), Article 22411. <https://doi.org/10.1038/s41598-024-72470-4>
- Schwammenthal, Y., Rabinowitz, T., Basel-Salmon, L., Tomashov-Matar, R., & Shomron, N. (2025). Noninvasive fetal genotyping using deep neural networks. *Briefings in Bioinformatics*, 26(1), Article bbaf067. <https://doi.org/10.1093/bib/bbaf067>
- Shear, M. A., Robinson, P. N., & Sparks, T. N. (2025). Fetal imaging, phenotyping, and genomic testing in modern prenatal diagnosis. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 98, Article 102575. <https://doi.org/10.1016/j.bpobgyn.2024.102575>
- Sunderajah, V., Guni, A., Liu, X., Collins, G. S., Karthikesalingam, A., Markar, S. R., Golub, R. M., Denniston, A. K., Shetty, S., Moher, D., Bossuyt, P. M., Darzi, A., Ashrafian, H., & STARD-AI Steering Committee. (2025). The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nature Medicine*, 31(10), 3283–3289. <https://doi.org/10.1038/s41591-025-03953-8>
- Sparks, T. N., & Chitty, L. S. (2026). The importance of clinical acumen for prenatal diagnosis in an increasingly technological world. *Prenatal Diagnosis*, 46(5–6), 619–622. <https://doi.org/10.1002/pd.70051>
- Stanley, K. E., Thienpont, B., & Vermeesch, J. R. (2025). Expanding the scope of non-invasive prenatal screening. *Nature Genetics*, 57(11), 2644–2654. <https://doi.org/10.1038/s41588-025-02377-x>
- Suhag, A., Kidd, J., McGath, M., Rajesh, R., Gelfinbein, J., Cacace, N., Monteleone, B., & Chavez, M. R. (2023). ChatGPT: A pioneering approach to complex prenatal differential diagnosis. *American Journal of Obstetrics & Gynecology MFM*, 5(8), Article 101029. <https://doi.org/10.1016/j.ajogmf.2023.101029>
- Tang, J., Han, J., Jiang, Y., Xue, J., Zhou, H., Hu, L., Chen, C., & Lu, L. (2023). An innovative three-stage model for prenatal genetic disorder detection based on region-of-interest in fetal ultrasound. *Bioengineering*, 10(7), Article 873. <https://doi.org/10.3390/bioengineering10070873>
- Tang, J., Han, J., Xue, J., Zhen, L., Yang, X., Pan, M., Hu, L., Li, R., Jiang, Y., Zhang, Y., Jing, X., Li, F., Chen, G., Zhang, K., Zhu, F., Liao, C., & Lu, L. (2023). A deep-learning-based method can detect both common and rare genetic disorders in fetal ultrasound. *Biomedicine*, 11(6), Article 1756. <https://doi.org/10.3390/biomedicine11061756>
- Tse, O. Y. O., Jiang, P., Cheng, S. H., Peng, W., Shang, H., Wong, J., Chan, S. L., Poon, L. C. Y., Leung, T. Y., Chan, K. C. A., Chiu, R. W. K., & Lo, Y. M. D. (2021). Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 118(5), Article e2019768118. <https://doi.org/10.1073/pnas.2019768118>
- Tseng, J.-J., Lu, C.-H., Li, J.-Z., Lai, H.-Y., Chen, M.-H., Cheng, F.-Y., & Kuo, C.-E. (2023). An open dataset of annotated metaphase cell images for chromosome identification. *Scientific Data*, 10(1), Article 104. <https://doi.org/10.1038/s41597-023-02003-7>
- Tsui, W. H. A., Ding, S. C., Jiang, P., & Lo, Y. M. D. (2025). Artificial intelligence and machine learning in cell-free-DNA-based diagnostics. *Genome Research*, 35(1), 1–19. <https://doi.org/10.1101/gr.278413.123>
- Turrieff, A. E., Annunziata, C. M., Malayeri, A. A., Redd, B., Pavelova, M., Goldlust, I. S., Rajagopal, P. S., Lin, J., & Bianchi, D. W. (2024). Prenatal cfDNA sequencing and incidental detection of maternal cancer. *The New England Journal of Medicine*, 391(22), 2123–2132. <https://doi.org/10.1056/NEJMoa2401029>
- Umans, E., Dewilde, K., Williams, H., Deprest, J., & Van den Bosch, T. (2024). Artificial intelligence in imaging in the first trimester of pregnancy: A systematic review. *Fetal Diagnosis and Therapy*, 51(4), 343–356. <https://doi.org/10.1159/000538243>
- Van den Veyver, I. B., Chandler, N., Wilkins-Haug, L. E., Wapner, R. J., Chitty, L. S., & ISPD Board of Directors. (2022). International Society for Prenatal Diagnosis updated position statement on the use of genome-wide sequencing for prenatal diagnosis. *Prenatal Diagnosis*, 42(6), 796–803. <https://doi.org/10.1002/pd.6157>
- Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., Denniston, A. K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B. A., Mathur, P., McCradden, M. D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D. S. W., . . . DECIDE-AI expert group. (2022). Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature Medicine*, 28(5), 924–933. <https://doi.org/10.1038/s41591-022-01772-9>
- Vestito, L., Jacobsen, J. O. B., Walker, S., Cipriani, V., Harris, N. L., Haendel, M. A., Mungall, C. J., Robinson, P., & Smedley, D. (2024). Efficient reinterpretation of rare disease cases using Exomiser. *npj Genomic Medicine*, 9(1), Article 65. <https://doi.org/10.1038/s41525-024-00456-2>
- World Health Organization. (2021). *Ethics and governance of artificial intelligence for health*. <https://www.who.int/publications/i/item/9789240029200>
- Wu, K.-H., Huang, H.-W., Lin, C. Y., Huang, H.-T., Fan, T.-Y., Chen, Y.-P., Chang, Y.-C., Hsu, T.-Y., & Lan, K.-C. (2026). Development of an artificial intelligence-based chromosome interpretation system for amniotic fluid karyotyping. *International Journal of Molecular Sciences*, 27(4), Article 1746. <https://doi.org/10.3390/ijms27041746>
- Yang, C., Li, T., Dong, Q., & Zhao, Y. (2023). Chromosome classification via deep learning and its application to patients with structural abnormalities of chromosomes. *Medical Engineering & Physics*, 121, Article 104064. <https://doi.org/10.1016/j.medengphy.2023.104064>
- Yang, J., Ding, X., & Zhu, W. (2018). Improving the calling of non-invasive prenatal testing on 13-/18-/21-trisomy by support vector machine discrimination. *PLOS ONE*, 13(12), Article e0207840. <https://doi.org/10.1371/journal.pone.0207840>
- Yu, S. C. Y., Jiang, P., Peng, W., Cheng, S. H., Cheung, Y. T. T., Tse, O. Y. O., Shang, H., Poon, L. C., Leung, T. Y., Chan, K. C. A., Chiu, R. W. K., & Lo, Y. M. D. (2021). Single-molecule sequencing reveals a large population of long cell-free DNA molecules in maternal plasma. *Proceedings of the National Academy of Sciences of the United States of America*, 118(50), Article e2114937118. <https://doi.org/10.1073/pnas.2114937118>

- Yuan, Y., Chai, X., Liu, N., Gu, B., Li, S., Gao, Y., Zhou, L., Liu, Q., Yang, F., Liu, J., Qiu, J., Zhang, J., Hou, Y., Cen, M., Tian, Z., Tang, W., Zhang, H., Chen, F., Yin, Y., & Wang, W. (2020). FF-QuantSC: Accurate quantification of fetal fraction by a neural network model. *Molecular Genetics & Genomic Medicine*, 8(6), Article e1232. <https://doi.org/10.1002/mgg3.1232>
- Zhang, J., Li, J., Saucier, J. B., Feng, Y., Jiang, Y., Sinson, J., McCombs, A. K., Schmitt, E. S., Peacock, S., Chen, S., Dai, H., Ge, X., Wang, G., Shaw, C. A., Mei, H., Breman, A., Xia, F., Yang, Y., Purgason, A., . . . Eng, C. M. (2019). Non-invasive prenatal sequencing for multiple Mendelian monogenic disorders using circulating cell-free fetal DNA. *Nature Medicine*, 25(3), 439–447. <https://doi.org/10.1038/s41591-018-0334-x>
- Zhang, L., Dong, D., Sun, Y., Hu, C., Sun, C., Wu, Q., & Tian, J. (2022). Development and validation of a deep learning model to screen for trisomy 21 during the first trimester from nuchal ultrasonographic images. *JAMA Network Open*, 5(6), Article e2217854. <https://doi.org/10.1001/jamanetworkopen.2022.17854>
- Zhong, W., Sun, M., Yao, S., Liu, Y., Peng, D., Liu, Y., Yang, K., Gao, H., Yan, H., Hao, W., Yan, Y., & Yin, C. (2025). Enhancing the accuracy of Human Phenotype Ontology identification: Comparative evaluation of multimodal large language models. *Journal of Medical Internet Research*, 27, Article e73233. <https://doi.org/10.2196/73233>
- Zhu, Y., Chau, M. H. K., Wang, H., Song, N., Wei, R., Suen, K. W., Chan, A. C. S., Hung, W. C., Cao, Y., Dong, Z., Leung, T. Y., Cheung, S. W., & Choy, K. W. (2025). Clinical validation of artificial intelligence-assisted karyotyping on peripheral blood in a cytogenetic diagnostic laboratory. *Human Genetics*, 144(11–12), 1269–1276. <https://doi.org/10.1007/s00439-025-02789-z>

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). This publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© Copyright (2026): Author(s). The licensee is the journal publisher. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
<https://pr.sdiarticle5.com/review-history/167077>