CellPress
OPEN ACCESS

Trends in
Molecular Medicine

Special Issue: Reproductive and Sexual Health

Feature Review

Data-Driven Modeling of Pregnancy-Related Complications

Camilo Espinosa,[1,2,4] Martin Becker,[1,2,4] Ivana Marić,[3,4] Ronald J. Wong,[3] Gary M. Shaw,[3] Brice Gaudilliere,[1,3] Nima Aghaeepour,[1,2,3] David K. Stevenson[3,*], the Prematurity Research Center at Stanford:[5]

Ina A. Stelzer, Laura S. Peterson, Alan L. Chang, Maria Xenochristou, Thanaphong Phongpreecha, Davide De Francesco, Michael Katz, Yair J. Blumenfeld, and Martin S. Angst

A healthy pregnancy depends on complex interrelated biological adaptations involving placentation, maternal immune responses, and hormonal homeostasis. Recent advances in high-throughput technologies have provided access to multiomics biological data that, combined with clinical and social data, can provide a deeper understanding of normal and abnormal pregnancies. Integration of these heterogeneous datasets using state-of-the-art machine-learning methods can enable the prediction of short- and long-term health trajectories for a mother and offspring and the development of treatments to prevent or minimize complications. We review advanced machine-learning methods that could: provide deeper biological insights into a pregnancy not yet unveiled by current methodologies; clarify the etiologies and heterogeneity of pathologies that affect a pregnancy; and suggest the best approaches to address disparities in outcomes affecting vulnerable populations.

Highlights

A multitude of clinical, biological, environmental, and demographic factors influence the trajectory of a pregnancy. Maternal genetics, environment, stress, nutrition, medical history, socioeconomic status, and racial and ethnic background all play a role in determining the success of a pregnancy.

Diverse data sources are available for the study of pregnancy and prediction of adverse outcomes, including electronic health records (EHRs) and administrative claims data, high-throughput multiomics data for characterizing biological systems, and more complex sources like time series, imaging and video data, and text.

Recent advances in multiview, multitask, and deep learning allow joint modeling across data sources as well as across outcomes and demonstrate the vast potential of such integrated approaches.

Complexities of Pregnancy

Pregnancy is a complex process in which each stage – implantation, decidualization, placentation, organogenesis, the establishment of the maternal–fetal interface, fetal growth, and, finally, parturition – depends on the successful completion of the previous stage [1]. The exact etiologies of most adverse pregnancy outcomes remain unknown, challenging the development of therapeutic interventions. Preterm birth (PTB), defined as a birth before 37 weeks of gestation, is the leading cause of global infant morbidity and mortality. Spontaneous PTB (sPTB) has many antecedents and has been attributed to various pathological mechanisms including intra-amniotic infection, disruption of maternal–fetal tolerance, vascular disorders, myometrial stretching, and cervical incompetence, thus appearing to be a syndrome with many causes [2]. Moreover, the current definition of sPTB, based solely on a gestational age (GA) of less than 37 weeks, is insufficient to precisely identify neonates who will develop adverse outcomes. Further challenging this definition is the inability to accurately estimate GA in low- and middle-income countries (LMICs) [3]. Similarly, there is growing evidence that the two main types of preeclampsia – early versus late onset – in addition to having different risk factors have different pathophysiologies [4]. Various adverse pregnancy outcomes – fetal growth restriction (FGR), PTB, preterm premature rupture of membranes (PPROM), late spontaneous abortion, and placental abruption – have been associated with a common culprit: an abnormal placenta [5,6]. These findings point to the complexity of these outcomes and of pregnancy itself.

In addition to – and perhaps as a consequence of – this inherent complexity of pregnancy, the corresponding adverse outcomes are influenced by various interrelated factors, including genetic and immunological predisposition as well as medical history and social determinants (e.g., race/ethnicity,

[1]Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Stanford, CA, USA
[2]Department of Biomedical Data Sciences, Stanford University, Stanford, CA, USA
[3]Department of Pediatrics, Division of Neonatal and Developmental Medicine, Stanford University School of Medicine, Stanford, CA, USA
[4]These authors contributed equally
[5]Secondary author list of consortium members (members of the Prematurity Research Center)

*Correspondence:
dstevenson@stanford.edu
(D.K. Stevenson).

socioeconomic status, education) [7]. New diseases and environmental factors – such as coronavirus disease 2019 (COVID-19) and climate change – can affect pregnancy in many complex ways [6,8,9]. Disparities in outcomes often seem to be related to racial, socioeconomic, and geographic factors [10]. However, the assessment of a woman's predisposition to a specific complication based on clinical, demographic, and social data – a key step – is limited in its predictive accuracy [11].

The recent availability of high-throughput, molecular-level data from genomic, transcriptomic, proteomic, metabolomic, and single-cell immunological measurements, together with advanced computational and statistical tools, has provided an opportunity for analyses of these comprehensive and detailed datasets and for the integration of biological and nonbiological biomarkers. Such integrated approaches can lead to more accurate inferences about the diversity and multiplicity of causes of pathologies related to pregnancy, providing more specific and generalizable signatures of pregnancy-related pathologies.

In this review, we provide an overview of major clinical challenges in pregnancy and the available data sources to study pregnancy, as well as state-of-the-art **machine-learning** (see Glossary) methodology to analyze these data. We close with an outlook on how a combination of diverse data sources and advanced machine learning can yield a holistic view of pregnancy.

Major Clinical Challenges in Pregnancy

The Healthy Progression of Pregnancy Depends on Complex Biological Adaptations

The establishment, maintenance, and completion of a healthy pregnancy depends on a series of interrelated biological adaptations. Derangements in the processes of placentation, maternal immune adaptation, and hormonal homeostasis, among others, contribute to the pathogenesis of adverse outcomes [12]. For example, during pregnancy the maternal immune system must maintain tolerance to the fetoplacental unit while still protecting the mother and fetus against invading pathogens. Failure to do so is strongly implicated in almost all adverse outcomes of pregnancy [12,13]. Similarly, pregnancy stages are orchestrated in part by specific chronological changes in the maternal endocrine system, and any dysregulation of hormonal homeostasis can lead to pathogenesis [14,15].

Pathogenic Processes in Pregnancy Lead to Complications

Pathogenic processes can interfere with the critical biological adaptations during pregnancy and lead to pregnancy-related complications. PTB, preeclampsia, and FGR arise from such processes and complicate 15–20% of pregnancies. While most of these disorders present clinically in the third trimester, treatment after the onset of clinical symptoms results in only mild to moderate improvements. Because implantation, uterine remodeling, and placental development occur early in pregnancy, identifying molecular changes weeks before the clinical onset of disease has the potential to dramatically improve our diagnostic and therapeutic capacities and decrease the associated maternal and neonatal morbidity and mortality.

Prematurity

Prematurity remains a leading cause of neonatal morbidity and mortality, and children born preterm often face lifelong challenges including permanent physical and neurodevelopmental disabilities [16]. PTB etiologies are often categorized into different subgroups [17,18]. Iatrogenic PTB is defined as a PTB in which physicians either induce labor or proceed with Cesarean delivery at early GAs in response to a severe underlying maternal and/or fetal complication. sPTB results from either PPROM or spontaneous preterm contractions/preterm labor (PTL). Although infection and inflammation have traditionally been thought to mediate a significant proportion of sPTBs, recent advances have also highlighted the role of pathological immune responses, maternal sociodemographic stressors (including bias and racism) [19,20], cervical

Glossary

Cardiotocography (CTG): recording of the fetal heartbeat and uterine contractions during pregnancy.
Complex data: data that cannot be directly represented in tabular form, such as text, images, videos, time series, or networks.
Deep learning: family of machine-learning algorithms based on neural networks.
Dimensionality reduction: transformation of data with a large number of observations and/or features to a meaningful representation with a smaller number of representative variables.
High-dimensional data: data with a large number of features (e.g., thousands or tens of thousands) that typically exceed the number of observations.
Machine learning: a wide range of statistical methods for making inferences about data given by a set of features (e.g., clinical measurements of patients).
Neural networks: collections of interconnected nodes organized in multiple layers, such that nodes in one layer are connected to nodes in its neighboring layers via weighted links. The learning in a neural network is performed by adjusting the weights to perform a task at hand with minimum observed error (i.e., with maximum accuracy).
Semi-supervised learning: a hybrid of supervised and unsupervised learning in which the outcome is known for only a limited number of samples.
Supervised learning: algorithms that learn to predict a certain property or an outcome (e.g., PTB) (also called a label, a target, a response variable, or an output) associated with a given set of features (also called input features, explanatory variables, or input).
Tabular data: data with a tabular structure where each row represents a sample and the columns correspond to categorical or numerical features.
Tocolytics: medications used to suppress premature labor.
Unsupervised learning: attempts to derive hidden structure from the data with no knowledge of an outcome.

disorders (e.g., cervical incompetence), and uterine overdistension [as is the case in multiple gestations or in polyhydramnios (excess amniotic fluid volume)]. Unfortunately, irrespective of the etiology, PTB remains extremely difficult to predict, partly because of heterogeneous and multifactorial underlying mechanisms and partly because clinicians lack the necessary tools to identify these mechanisms prior to the onset of clinical symptoms. Moreover, treatment of PTL and/or PPROM using **tocolytics**, antibiotics, or other measures has only limited benefit in delaying delivery once clinical symptoms have begun [21].

Preeclampsia

Preeclampsia is a multiorgan disorder complicating 2–8% of pregnancies [22]. Although the hallmarks of preeclampsia are new-onset maternal hypertension and proteinuria, preeclampsia can have significant adverse effects on a variety of organs including the central nervous system, lungs, liver, kidneys, and heart, all mediated via generalized vasoconstriction [22]. Preeclampsia may present as early- or late-onset preeclampsia, with the early-onset form indicative of more severe disease and with increased rates of complications [23]. While the clinical manifestations of preeclampsia occur most often in the third trimester, it is hypothesized that the pathophysiology of preeclampsia begins in the first trimester. Specifically, abnormal invasion of the placental trophoblasts into the uteroplacental interface is believed to result in maldevelopment of the spiral arteries and hence lead to inadequate perfusion of the placenta [24]. This early-onset abnormal placentation cascade can also have direct third trimester ramifications as preeclampsia can result in placental abruption, PTB, and/or FGR. The adverse effects of preeclampsia are likely to extend beyond maternal and neonatal morbidity. Recent studies have begun to elucidate a link between preeclampsia and long-term maternal and offspring cardiovascular disease, likely to occur via direct and permanent cardiac remodeling [25,26]. Therefore, the identification of novel predictive tools, which allow timely interventions, may ultimately lead not only to a reduction in maternal and neonatal morbidity but also to improvements in the long-term health of women.

FGR

Although FGR is often defined as an estimated fetal weight less than the tenth percentile [27], most cases are not pathological but rather 'constitutional', based on maternal or paternal size. Pathological FGR complicates 3–5% of pregnancies and results from abnormal genetics, infections [e.g., cytomegalovirus (CMV)], anatomical malformations, placental pathologies, and/or underlying maternal disorders leading to decreased placental perfusion [28]. The long-term neonatal and pediatric effects of FGR have been well characterized and include physical, metabolic, and neurodevelopmental challenges [27]. While most FGR is diagnosed in the third trimester, predicting FGR in the first or second trimester is difficult, and there are no proven therapies to alter fetal growth *in utero*. Interestingly, there is a link between FGR and metabolic syndrome in the offspring, highlighting a vicious cycle in which a pathological pregnancy adversely affects the long-term health and possible eventual pregnancy of the offspring [29].

Stillbirth

Stillbirth – fetal death after 20 weeks of gestation – remains a significant global health burden in both LMICs and high-income countries (HICs) [30,31]. While the cause of intrauterine fetal death is multifactorial, common complications that can lead to stillbirth include fetal congenital anomalies, placental pathologies, maternal health disorders, and obstetric syndromes [32]. Furthermore, a pregnancy that ends with a stillbirth will have long-term physical and psychological consequences on the mother's health and may endanger the success of future pregnancies [33,34].

Clinical and Social Determinants of Maternal and Neonatal Health

Clinical and Administrative Data Are Good Starting Points

Various population-level data sources have been used to identify factors that impact maternal and offspring health. One of the main such sources – electronic health records (EHRs) – includes demographics, anthropometrics, diagnoses, vital signs, medications, and various clinical measurements, offering a detailed look into a patient's medical history. Administrative insurance claims data, containing large sample sizes over large geographic areas but with less granularity, have been successfully mined to create and analyze pregnancy cohorts [35–37]. Studies involving EHR-based models for the prediction of gestational diabetes mellitus (GDM), preeclampsia, and PTB illustrate how such data sources are relevant in the context of reproductive health [38–40]. EHRs can also be used to characterize the currently unknown pharmacological effects that a broad range of drugs might have on the physiology of pregnancy [41,42]. Well-established measures of health such as hematology panels and urinalyses continue to provide useful information when assessing maternal and fetal health. For instance, higher neutrophil counts and neutrophil-to-lymphocyte ratios – both inflammatory markers – have been associated with the development of GDM and decreased fetal growth [43]. Additionally, proteinuria and bacteriuria, determined through urinalysis and urine culture, respectively, are important predictors of life-threatening complications like preeclampsia [44,45]. Advances in artificial intelligence (AI) have recently enabled clinicians to use some of these traditional, low-cost tests in more sophisticated predictive and diagnostic models than previously possible [46].

Lifestyle Choices and Social Determinants Impact Maternal and Fetal Health

Recent results have revealed that psychological, nutritional, and other lifestyle-related factors can affect the course of a pregnancy. Stress – broadly defined using clinical diagnoses of depression and anxiety and physiological measurements such as elevated blood pressure – can affect fetal development with consequences that extend into adulthood [47,48]. Increased levels of stress have been associated with preeclampsia and PTB through poorly understood biological mechanisms [49–51]. The influence of maternal nutrition on fetal development is better characterized [52]. Both malnutrition and obesity affect the adapted maternal metabolism in pregnancy and correlate with infertility, PTB, and FGR [53,54]. Similarly, micronutrient deficiencies, particularly vitamin D, are epidemiologically linked to preeclampsia, GDM, and developmental fetal defects, demonstrating the importance of maternal nutrition during gestation [55–58].

Racial and ethnic background, socioeconomic status, and living environment act via complex mechanisms that make measurements of their direct effects on maternal and fetal health difficult [59]. For example, a recent study found two distinct pathways by which stress can lead to complications during pregnancy demarcated by racial and ethnic lines, such that Hispanic and African-American women progress towards the same adverse outcomes fueled by similar stressors, but in biologically different ways [60]. Lower socioeconomic status and levels of education decrease access to prenatal care (even when costs are not prohibitive) to the detriment of the mother and her offspring [61,62]. Exposures to air pollution, heavy metals, and toxic compounds also often distribute across socioeconomic and geographic lines and have been associated with increased risks of developmental defects and other adverse outcomes [63–67].

Machine learning based on clinical and social determinants to model adverse pregnancy outcomes yields promising results that – dependent on the data used – can enhance our ability to predict adverse outcomes and enable deeper insights into the underlying processes of pregnancy. However, they have not yet been shown to yield predictive models of adverse pregnancy outcomes such as PTB in a generalizable and robust manner, making biological profiling of pregnancy a major priority.

Biological Profiling of Maternal and Neonatal Health

Recent advances in high-throughput technologies have provided access to multidomain or multiomic biological data, enabling studies to understand pregnancy from the viewpoints of genetics and epigenetics as well as a wide range of other omics. For such **high-dimensional data**, characterized by a very large number (e.g., tens of thousands) of measurements obtained for a small number of patients, machine-learning approaches are a necessary analytical tool.

Genetics and Epigenetics for Characterizing Pregnancy

Pregnancy occurs against the genetic background and epigenetic programming of both the mother and the fetus. Maternal genetic variants affect the success of essential developmental processes such as decidualization and placentation, whereas both maternal and fetal genetics play a role in maintaining the delicate immune balance that characterizes embryonic and fetal development [68–70]. While genetic contributions to different complications in pregnancy are probably polygenic, genome-wide association studies (GWASs) have found genetic variants associated specifically with sPTB, illustrating how a mother's genetic makeup can set the stage for fetal development and pregnancy outcome even before conception [71–73]. By contrast, maternal epigenetic modifications reflect the state of different cell types across tissues as they reprogram to prepare for fetal growth [74,75]. Quantification of these modifications – by measuring DNA methylation, histone markers, and chromatin accessibility in the cell subsets of interest – offers insights into whether different cell types in the maternal–fetal interface are adapting properly for a successful pregnancy [76,77].

Omics Data Analyses Provide a Deep View of the Biological Pathways of Pregnancy

A more complete characterization of the mother's biological state can be obtained by utilizing, in addition to genetic and epigenetic data, high-throughput methods to assess the maternal transcriptome, metabolome, proteome, and microbiome (Box 1). Given the massive pregnancy-induced changes in maternal tissue-specific gene expression and the accompanying metabolic, hormonal, and immune adaptations, these modalities can reveal different facets of maternal and fetal health at specific points in time [78–82]. For example, cell-free RNA (cfRNA) from blood or amniotic fluid can be used to monitor the dynamics of fetal development and placentation, with implications for the early diagnosis of preeclampsia by the detection of biomarkers of dysfunctional placentation in maternal plasma [83,84]. Microbiomics can be used to survey the vaginal, oral, and gut microbiomes, which play key roles in maintaining the maternal immune balance necessary to tolerate the fetus [85]. Dysbiosis of the vaginal microbiome, such as is seen in bacterial vaginosis, can lead to PTB through inflammatory processes or intra-amniotic infections originating from vaginal microbes [86]. Similarly, pathogenic oral microbes can lead to PTB via systemic inflammation from oral disease or through hematogenous colonization of the amniotic fluid [87]. Finally, the gut microbiome, with its role in the metabolism of small molecules and hormones and in the regulation of metabolic and immune processes, has been linked to GDM and PTB when unbalanced [88]. Overall, the various omics technologies hold great potential to help elucidate the biological components of the pathologies of pregnancy.

Machine Learning for Modeling Adverse Pregnancy Outcomes

Machine Learning

Machine learning encompasses a wide range of statistical methods used to make inferences about data given by a set of features (e.g., clinical measurements of patients) [89]. It includes **supervised learning** – algorithms that learn to predict a certain property or an outcome (e.g., PTB) (also called a label, a target, a response variable, or an output) associated with a given set of features (also called input features, explanatory variables, or input). It is supervised because the outcome is known for a large enough data sample (i.e., number of patients) allowing a relationship between

Box 1. Multiomics Phenotyping for Systems Biology

Systems biology approaches have revealed the complexity of the biological components whose mutual interactions maintain homeostasis in a healthy individual. Cell expression profiles, metabolites, proteins, and the microbiota around the body all play roles in homeostasis and the dysregulation of their interactions can give rise to disease. Recent high-throughput technologies have enabled the quantification of these biological components with increased resolution, giving rise to omics data.

Transcriptomics

Transcriptomic data faithfully track tissue-specific cellular programs through the sequencing of different types of RNA molecules. Sequencing technologies have enabled the quantification of both protein-coding mRNA and regulatory miRNAs, circular (circ)RNAs, and long noncoding (lnc)RNAs [176]. cfRNA from blood samples has also emerged as a promising application of transcriptomic profiling in the context of noninvasive biomarkers for use in clinical practice [177]. The limitations in cell-free transcriptomic data lie in the constraints of sequencing technology and our inability to determine the exact tissue of origin of the RNA sequenced.

Metabolomics

Metabolomics provides a comprehensive view of the metabolites – specifically, small water-soluble bioactive molecules – that are present in a biological system. Metabolites in a sample are quantified using NMR and mass spectrometry (MS), which allows measurements of high precision and sensitivity. However, the large number of unannotated metabolites and the difficulty of establishing their biological function are challenges that prevent the full exploitation of this omic [178].

Proteomics

Measuring the proteome provides insight into the biological functions occurring in an individual at a given point in time by quantifying the abundance of proteins of interest. Methods used for proteomics include MS and antibody- or aptamer-based profiling, with the latter two necessitating previous knowledge of the proteins to be quantified. Nonetheless, the large variations in the relative abundances of proteins in a sample pose a challenge to proteomics analysis, and the cost of this omic can be prohibitive [179].

Microbiomics

The microbiome encompasses tissue-resident commensal organisms and is known to critically contribute to the modulation of human physiology. Next-generation sequencing (NGS) is used to profile 16S ribosomal RNA or, together with shotgun metagenomics, to unravel the microbial landscape in a tissue of interest. As with transcriptomics, the constraints of sequencing technologies can limit the detection of rare variants that may be significant when studying the relation of the microbiome and a relevant phenotype [180].

the features and the outcome to be learned and a prediction model to be developed. By contrast, **unsupervised learning** has no knowledge of the outcome and attempts to derive hidden structure from the data [90]. **Semi-supervised learning** is a hybrid of supervised and unsupervised learning in which the outcome is known for only a limited number of samples. In the context of pregnancy, we mainly focus only on supervised learning in which features such as medical history or measurements from an omics assay (input) are used to predict outcomes such GA or PTB (output). Most commonly, the input is represented in a structured, tabular form where each row represents a sample and the columns correspond to categorical or numerical features, and the output is a single outcome. For more information about supervised machine learning as well as general evaluation techniques, please refer to Box 2. For a discussion of different data types and their properties and challenges see Box 3.

Machine-Learning Models for Adverse Pregnancy Outcomes

Machine learning is well suited for predictive modeling of pregnancy outcomes [91,92] and is becoming more prevalent [93–105] due to its ability to model highly complex relationships between measured features and outcomes. The majority of previous work focused on modeling techniques that incorporate one or two data sources, including clinical [95] as well as derived numerical data from another source such as blood samples [39], Doppler ultrasound, echosonography, or magnetic resonance imaging (MRI) readings [101], or mental health assessments [106]. Together, these datasets are combined into a single, structured table of samples and features. See Box 3 for more information on different data types and properties.

Box 2. Machine Learning: Supervised Learning

In supervised learning, machine-learning models learn to predict certain properties or an outcome (also called labels, targets, response variables, or output) associated with a sample represented by a set of features (also called input features, explanatory variables, or input) [89]. To achieve this, a machine-learning model assumes a certain input/output relationship (e.g., linear) and is then trained based on a set of samples (i.e., training data) for which both input and output are known. This training procedure typically comprises iterative adjustment of the parameters of the chosen input/output relationship. After training, the model can then be applied to previously unseen samples and can make predictions about the desired output. Different machine-learning models differ in which types of input and output they can process as well as how well they can cope with the different complexities of the relationship between the given input and output. Most commonly, the input is represented in tabular form where rows represent samples and columns correspond to categorical or numerical features. Depending on the class of functions used to model the input/output relationship, various machine-learning model classes exist, including linear regression variants [181], rule- and decision-tree-based systems [182], ensemble methods like random forests [183] or gradient-boosted trees [184], support vector machines [185], nearest-neighbors approaches [186], and Bayesian techniques [187]. Artificial neural networks and deep learning [188] can also be applied to tabular data, although in this task other methods often outperform fully connected neural networks [189]. For further discussion on different input types and their properties, see Box 3. The evaluation of supervised machine-learning models is based on datasets where the input and output are known. Based on this, the most common approach is cross-validation, where this dataset is split into pairs of a training and a test set where the test sets do not overlap. For each split, the model is trained and optimized on the training set while the predictive power of the trained model is assessed on the test set. The final score is an average score over all splits.

Computational Modeling of Pregnancy Outcomes Using High-Throughput Biological Data

High-throughput multiomics data poses two unique challenges for state-of-the-art machine-learning models and advanced statistical tools: (i) typically, the large number of features (high dimensionality) and small number of samples (e.g., patients) of omics data; and (ii) the heterogeneity of the different omics to be integrated – a challenging task even without considering high dimensionality.

The analysis of high-dimensional data with small numbers of samples is commonly enabled by two approaches: (i) reduction of the number of features; and (ii) sparse modeling. Reduction of the number of features is usually applied as a preprocessing step. It reduces the number of features either by feature selection [107], which filters out redundant features, or by **dimensionality reduction**, which compresses the information into a representation containing a smaller number of representative features [108] and is done in an unsupervised manner. A common example of dimensionality reduction is principal component analysis (PCA). It reduces noise and simplifies the data and thus is often applied for visualization. The second approach, sparse modeling [109], directly models an outcome in a supervised manner. It extracts the most informative features

Box 3. Data and Its Properties

This review distinguishes between tabular and more complex data. Here, tabular data is referred to as data with a tabular structure where each row represents a sample and the columns correspond to categorical or numerical features. However, such data can have very different properties. For example, small numbers of clinical variables may be available for large cohorts, while readings from transcriptomics assays provide many features that are available only for small cohorts (see tables in Figure 1 in the main text). Furthermore, the data themselves may differ: clinical data might contain a mix of categorical and numerical values; omics assays contain only numerical values; and the transcriptome is inherently sparse due to the fact that many genes are generally not active. Thus, while their structures are similar, the individual characteristics of different data sources pose unique challenges to machine-learning models analyzing this data and may require different algorithms (Box 2). This heterogeneity in data complicates their integration in multiomics settings, as discussed in the subsection 'Computational Modeling of Pregnancy Outcomes Using High-Throughput Biological Data'.

The tabular form forces the input to be limited to a fixed set of categorical or numerical features. However, machine-learning research has made tremendous progress on more complex data types such as text, images, videos, time series, and even complex networks, which this review refers to as complex data. In medical settings, such modalities (e.g., ultrasound images [146,151,152]) are often converted into handcrafted features to fit the paradigm of tabular data. Due to this inherent handcrafting process, these features are susceptible to human bias and information loss. In other domains, however, deep-learning methods have been shown to be highly versatile in directly processing complex data modalities without laborious and handcrafted preprocessing procedures. This enables novel avenues to take advantage of routinely collected but hardly analyzed modalities, as further discussed in the section 'Challenges and Potential of Complex Data'.

for that outcome and includes them in a prediction model that is often easy to interpret and efficient to compute. Sparse modeling also enables the integration of any additional *a priori* knowledge to guide and further sparsify the models [110]. Examples of the application of these techniques in the context of pregnancy include predicting GA by modeling the temporal dynamics of the proteome and immunome over the course of pregnancy [81,82]. Similarly, prediction models for preeclampsia risk have been developed from transcriptomic, proteomic, or metabolomic datasets [4,111], for PTB from immune, transcriptomic, and growth-related molecular factors [112], and for models of complications like macrosomia, FGR, GDM, or preeclampsia from cfRNA [113].

While coping with the heterogeneity of different omics is challenging, their integration has been demonstrated to improve predictive power. This is illustrated by a study on modeling GA [114] using stacked generalization [115] to integrate biological signals across seven different omics. In another study, the incorporation of transcriptomic and epigenetic data was shown to increase performance in the identification of GDM in addition to helping in the elucidation of its biological basis [116]. However, the individual characteristics of different omics as well as their inherent differences in numbers of features (e.g., sparse microbiome data versus targeted proteomics assays) pose unique challenges for their integration into statistical models. In particular, sparse data with many undefined values require different models than data with dense input. Another problem is that omics of high dimensionality but with low information content may preclude inclusion in the model of information from smaller, more dense omics. To address these issues, a wide variety of methods for multiomics integration is currently being developed in the field of bioinformatics [117–120], and there are community-driven efforts to maintain an overview of relevant work and software packages[i]. This work incorporates approaches from diverse fields of machine learning including Bayesian concepts [121,122], network analysis [117,118], and **deep-learning** techniques [123–125]. Similar to the single-omics case, methods for predictive studies often incorporate dimensionality reduction and data integration using a mixture of variable selection (e.g., [120]) and representation learning (e.g., [126]) to reduce the number of features and to calibrate their influence on the model between omics.

Finally, the use of single-cell immune profiling technologies for pregnancy research is still developing. There, one general limitation is the manual process of identifying homogeneous cell populations ('gating') for feature extraction from single-cell data, which disregards the detailed picture that single-cell technologies can provide. However, methods have recently emerged that integrate single-cell processing directly into the modeling pipeline [127,128].

In summary, pregnancy involves a number of intricate biological processes that can be better characterized by using a coordinated set of omics assays, typically available for a small set of patients. Although a variety of methods exist, the integration of such multiomics datasets is not yet broadly developed in the context of pregnancy research and requires the adaptation of existing and development of new machine-learning pipelines. This will allow the full exploitation of these novel data sources and may unveil new biological insights into pregnancy, the identification of risk factors, and the development of diagnostic predictive tests.

Challenges and Potential of Complex Data

Some of the valuable data sources for the modeling of pregnancy are available only in **complex data** formats such as time-series measurements (e.g., fetal heart rate or actigraphy), imaging data (e.g., ultrasound), and free text (e.g., diagnosis records, patient narratives). While these complex datasets make machine-learning modeling more challenging and have been less analyzed, they provide information critical for a more complete understanding of maternal and neonatal health (see Box 3 for a comparison of **tabular data** and complex data).

Time-Series Data Are a Promising Complex Modality

Clinical time-series data are essential for monitoring patients and health assessments in real time [e.g., in the intensive care unit (ICU)] as well as in retrospective settings (e.g., for long-term monitoring and diagnoses). One example in the context of pregnancy is fetal heart rate analysis to identify pathological fetal conditions like perinatal hypoxia, FGR, or fetal arrhythmias and heart anomalies [129]. Here, a common practice is to extract various quantitative parameters associated with fetal conditions such as short-term variability (STV) and long-term irregularity (LTI) of fetal heart rate measurements, which are then directly interpreted or converted into a tabular format that can be used for machine learning [129,130]. However, there is a wide variety of advanced machine-learning methods for time-series analyses and classifications, specifically deep learning and recurrent **neural networks**, that have the potential to directly model and predict pathological fetal conditions [131–133]. For example, deep learning can be directly applied to **cardiotocography (CTG)** recordings to predict fetal acidemia without handcrafted feature extraction akin to STV and LTI markers [134]. Instead, deep-learning architectures are built to process the given data directly, possibly identifying markers that may not have been known to practitioners.

The use of sensors and actigraphy to monitor sleep quality, activity patterns, and movement in patients is a novel approach to the quantification of lifestyle-related behaviors that might affect health [135,136]. Although this strategy has been underexplored, altered sleep patterns during pregnancy have been shown to lead to differential gene expression in mothers [137]. For this, machine-learning approaches for time-series data can be used to analyze smart-watch and actigraphy data [138], which can reveal insights into various pregnancy-related parameters [139]. Furthermore, in combination with the ubiquity of smartphones and wearable devices, a wide variety of activity recognition and semantic behavior analysis methods [140–143] can now be captured to obtain an even more detailed picture of potential behavioral risk factors.

Imaging Technologies for Biological Analyses

Another area where machine and specifically deep learning have shown great potential is image and video analyses [144–146]. However, in pregnancy-related research the corresponding methodology is still underrepresented although imaging technologies are routinely used to assess fetal health [146], and basic science as well as clinical research is increasingly adopting multiplexed imaging methods [147]. A prominent example is ultrasound-related modalities, considered a cornerstone of the clinical assessment of fetal health [148]. In particular, Doppler ultrasound – which can quantify fetal and umbilical blood flow – has been studied in the context of predicting adverse pregnancy outcomes in the first and second trimesters [149,150]. However, images and videos are often converted into (handcrafted) features for manual inspection or for input into machine-learning models using various feature-extraction methods [146,151,152]. While there are approaches to aid in the interpretation of pregnancy-related imaging data [146], only limited work has tried to tap into the full potential of image and video material to model and predict pregnancy-related outcomes [153].

Text-Based Information to Leverage Large-Scale Information Sources

Due to the rapid development of natural language processing tools, unstructured text is becoming a valuable information source. Application scenarios range from processing the text-based components of EHRs and patient narratives [154–156] to extracting (biological) knowledge [157] from previous research articles [158–161]. For example, EHRs can be used to automatically extract features based on computational word representations (word embeddings) for better prediction of adverse outcomes [38]. Similarly, machine learning can be used to extract and process subjective information using sentiment analysis [162,163] (e.g., to analyze discharge notes where sentiments have been shown to be associated with readmission and mortality rates [164]). Similar

Clinician's Corner

Major complications that may arise in pregnancy share overlapping etiologies and presentations and can lead to increased complexity in diagnosis and effective therapeutic intervention. Moreover, these diseases have both short- and long-term repercussions on the health of the mother and her offspring.

Integrating novel high-throughput biological profiling technologies with clinical and social determinants of maternal and fetal health will enable clinicians to describe the current state and possible risks of a pregnancy with greater accuracy.

Advanced machine-learning methods applied to the multitude of data sources now available to study pregnancy will reveal more robust biomarkers and specific disease signatures with applications to clinical practice.

Better biomarkers will permit earlier diagnosis of possible complications and may act as markers of therapeutic efficacy for clinicians to modulate treatment.

Specific and generalizable disease signatures will reveal how clinical, biological, environmental, and demographic factors contribute to the different pathologies of pregnancy, allowing better risk stratification of patients and an improved understanding of the pathogenesis of these diseases.

technology may be potentially used to monitor patient interactions for wellbeing during pregnancy. Overall, the processing of text-based information may open novel avenues to automatically capture and analyze information that either is not accessible without laborious examination of textual content or cannot be directly captured by biomarkers.

A Holistic View on Pregnancy

Recent advances in machine learning, including multimodal learning, multiview representation learning [165–171], and multitask learning [172,173], provide a unique opportunity for in-depth modeling of pregnancy and its pathologies (see Clinician's Corner). These areas of research aim to combine datasets from various modalities and across different tasks (e.g., prediction of outcomes) to develop an integrated model (Boxes 3 and 4). Multimodal models have been previously applied in biological data integration [169] and are particularly relevant for multiomics analysis and integration [118,126]. They can also combine regular tabular data, time series, images, and text into a joint holistic model for a multitude of predictive settings [166,167]. At the same time, pregnancy outcomes are highly interrelated and may point towards different phenotypes with similar pathologies. This interrelatedness can be exploited by the use of multitask learning, which takes advantage of information contained in related outcomes to make models more robust by preferring solutions that share common (e.g., biological) structures across these outcomes [173]. Deep learning has recently contributed significantly to advancing the field of multitask learning [133,172]. The combination of multimodal learning with multitask approaches [174] allows condensed representations of the inputs and modeled phenotypes (outcomes), which may lead to a novel holistic understanding of the underlying processes of pregnancy.

Concluding Remarks

Studies discussed in this review involve the integration of multiple types of biological, social, and clinical measurements. They are evidence of the importance of a holistic approach when attempting to characterize the state of maternal and fetal health and to model adverse outcomes (Figure 1). The discussed recent advances in machine learning are the foundation for such a holistic approach. These novel techniques may enable the discovery of more robust biomarkers and more specific disease signatures of pregnancy-related pathologies. Better biomarkers will allow earlier diagnosis of complications and improved evaluation of treatment efficacy [175]. Additionally, specific disease signatures will reveal actionable therapeutic targets to prevent complications. Ultimately, the synthesis of these diverse modalities of information about a mother and her offspring using state-of-the-art machine-learning methods will enable us to accurately describe their short- and long-term health trajectories. Reimagining our perspective of pregnancy

Box 4. Machine Learning: Multimodal and Multitask Learning for Holistic Models

As illustrated in Figure 1 in the main text, pregnancy is influenced by many factors that can be measured and tracked using various approaches. This yields highly multimodal data ranging from tabular data like clinical variables or biological assays to more complex data like ultrasound imaging or patient narratives. Multimodal machine learning, mostly enabled by deep learning, is a branch of machine learning that strives to integrate various data modalities into a single model. For this, it combines models that excel at processing each of these modalities into a single model. For example, an integration of histological images and gene expression data has been shown to improve pan-cancer prognosis [190]. At the same time, pregnancy outcomes are highly interrelated and may point towards different phenotypes with similar pathologies. This interrelatedness can be exploited by the use of multitask learning. Instead of predicting just one outcome like PTB or preeclampsia, multitask models predict several outcomes simultaneously. Thus, if configured correctly, the model has to learn the common underlying processes of these outcomes to achieve high predictive performance. This not only allows us to take advantage of information contained in related outcomes and make models more robust by preferring solutions that share common (e.g., biological) processes across these outcomes [173], but may also lead to novel insights into these common processes by interpreting the learned relations. Finally, multimodal and multitask models combine these two concepts and use several data modalities to simultaneously predict multiple tasks [174], and may open new avenues to build holistic models and a better understanding of pregnancy (Figure 1).

Outstanding Questions

How can state-of-the-art machine-learning methods be used to integrate available biological, clinical, and social data in the best way that can characterize pregnancy and predict pregnancy adverse outcomes?

Should definitions of PTB and preeclampsia phenotypes (early versus late onset) be made more precise to include more than just GA?

Which data modalities will serve as the most accurate source of information to model pregnancy and how can easily accessible data sources replace possibly more accurate expensive data sources?

How can a systems approach based on machine learning improve our understanding of pregnancy and adverse outcomes?

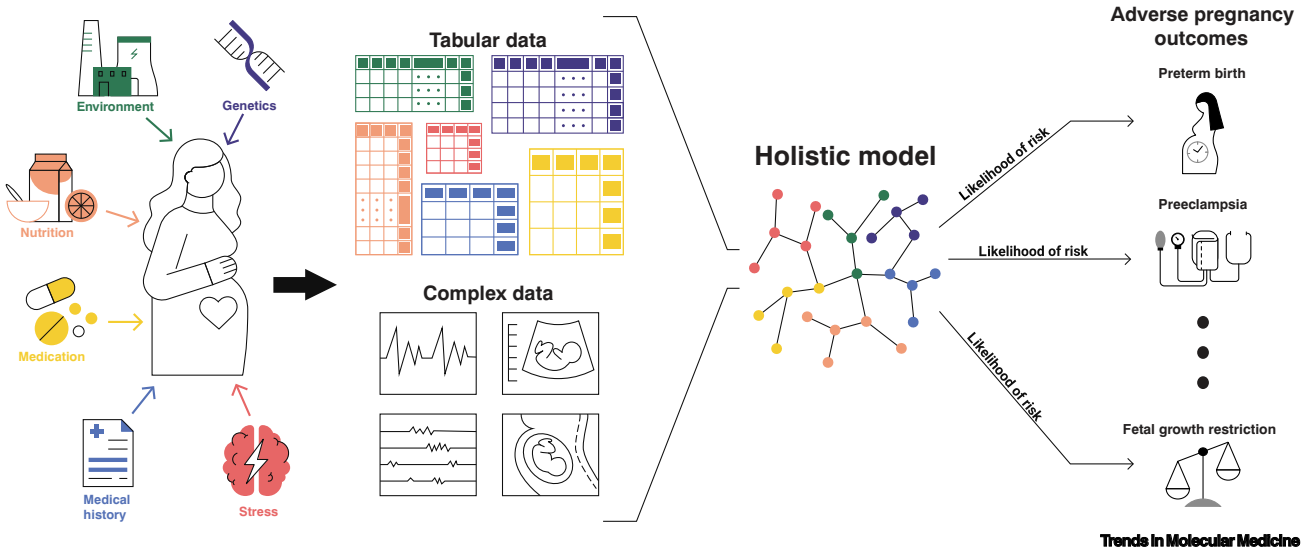

Figure 1. Incorporating Diverse Data Modalities to Build Holistic Models of Pregnancy Biology. The various factors that influence maternal and fetal health during gestation are measured to generate diverse, intercorrelated types of data. Machine-learning methods can be used to develop holistic models of maternal and fetal biology that capture the complex interactions between these modalities, reveal mechanistic insight into various adverse outcomes, and assist in diagnostics, therapeutics, and the generation of predictive analytics.

through this multimodal lens will provide relevant biological insight undetectable by current methodologies, further clarify the etiologies and explain the heterogeneity of the presentations of the pathologies that affect pregnant women, and shed light on the best approaches to address the disparities in outcomes that affect vulnerable populations (see Outstanding Questions).

Acknowledgments

We thank Anastasia Yarmolinskaya for her contribution to the illustration of Figure 1. This work was supported by the March of Dimes Prematurity Research Center at Stanford University, the Bill and Melinda Gates Foundation (OPP1112382, OPP1189911, and OPP1113682), the National Institutes of Health (R01AG058417 and R35 GM138353), the Burroughs Wellcome Fund (1019816), the Christopher Hess Research Fund, and the Robertson Foundation.

Declaration of Interests

No interests are declared.

Resources

[i]https://github.com/mikelove/awesome-multi-omics

References

1. Cha, J. *et al.* (2012) Mechanisms of implantation: strategies for successful pregnancy. *Nat. Med.* 18, 1754–1767
2. Romero, R. *et al.* (2014) Preterm labor: one syndrome, many causes. *Science* 345, 760–765
3. Rosenberg, R.E. *et al.* (2009) Determining gestational age in a low-resource setting: validity of last menstrual period. *J. Health Popul. Nutr.* 27, 332–338
4. Tarca, A.L. *et al.* (2019) The prediction of early preeclampsia: results from a longitudinal proteomics study. *PLoS One* 14, e0217273
5. Ilekis, J.V. *et al.* (2016) Placental origins of adverse pregnancy outcomes: potential molecular targets: an Executive Workshop Summary of the Eunice Kennedy Shriver National Institute of Child Health and Human Development. *Am. J. Obstet. Gynecol.* 215, S1–S46
6. Khalil, A. *et al.* (2020) Change in the incidence of stillbirth and preterm delivery during the COVID-19 pandemic. *JAMA* 324, 705–706
7. Stevenson, D.K. *et al.* (2020) Towards personalized medicine in maternal and child health: integrating biologic and social determinants. *Pediatr. Res.* Published online May 26, 2020. https://doi.org/10.1038/s41390-020-0981-8
8. Bekkar, B. *et al.* (2020) Association of air pollution and heat exposure with preterm birth, low birth weight, and stillbirth in the US: a systematic review. *JAMA Netw. Open* 3, e208243
9. Rylander, C. *et al.* (2013) Climate change and the potential effects on maternal and pregnancy outcomes: an assessment of the most vulnerable – the mother, fetus, and newborn child. *Glob. Health Action* 6, 19538

10. Stevenson, D.K. *et al.* (2019) Understanding health disparities. *J. Perinatol.* 39, 354–358
11. Poon, L.C. and Nicolaides, K.H. (2014) Early prediction of preeclampsia. *Obstet. Gynecol. Int.* 2014, 297397
12. Peterson, L.S. *et al.* (2020) Multiomic immune clockworks of pregnancy. *Semin. Immunopathol.* 42, 397–412
13. Han, X. *et al.* (2019) Differential dynamics of the maternal immune system in healthy pregnancy and preeclampsia. *Front. Immunol.* 10, 1305
14. Stelzer, I.A. and Arck, P.C. (2016) Immunity and the endocrine system. In *Encyclopedia of Immunobiology*, pp. 73–85, Elsevier
15. Solano, M.E. and Arck, P.C. (2019) Steroids, pregnancy and fetal development. *Front. Immunol.* 10, 3017
16. Martin, J.A. *et al.* (2019) Births: final data for 2018. *Natl. Vital Stat. Rep.* 68, 1–47
17. Goldenberg, R.L. *et al.* (2008) Epidemiology and causes of preterm birth. *Lancet* 371, 75–84
18. Romero, R. *et al.* (2006) The preterm parturition syndrome. *BJOG* 113, 17–42
19. Traylor, C.S. *et al.* (2020) Effects of psychological stress on adverse pregnancy outcomes and non-pharmacologic approaches for reduction: an expert review. *Am. J. Obstet. Gynecol. MFM* 2, 100229
20. Krieger, N. *et al.* (2020) Structural racism, historical redlining, and risk of preterm birth in New York City, 2013–2017. *Am. J. Public Health* 110, 1046–1053
21. American College of Obstetricians and Gynecologists' Committee on Practice Bulletins – Obstetrics (2016) Practice bulletin no. 171: management of preterm labor. *Obstet. Gynecol.* 128, e155–e164
22. American College of Obstetricians and Gynecologists' Committee on Practice Bulletins – Obstetrics (2020) Gestational hypertension and preeclampsia: ACOG practice bulletin, number 222. *Obstet. Gynecol.* 135, e237–e260
23. Wójtowicz, A. *et al.* (2019) Early- and late-onset preeclampsia: a comprehensive cohort study of laboratory and clinical findings according to the New ISHHP Criteria. *Int. J. Hypertens.* 2019, 4108271
24. Chaiworapongsa, T. *et al.* (2014) Pre-eclampsia part 1: current understanding of its pathophysiology. *Nat. Rev. Nephrol.* 10, 466–480
25. Andraweera, P.H. and Lassi, Z.S. (2019) Cardiovascular risk factors in offspring of preeclamptic pregnancies – systematic review and meta-analysis. *J. Pediatr.* 208, 104–113.e6
26. Yilgwan, C.S. *et al.* (2020) Profile of congenital heart disease in infants born following exposure to preeclampsia. *PLoS One* 15, e0229987
27. Society for Maternal–Fetal Medicine *et al.* (2020) Society for Maternal–Fetal Medicine Consult Series #52: diagnosis and management of fetal growth restriction: (replaces Clinical Guideline Number 3, April 2012). *Am. J. Obstet. Gynecol.* 223, B2–B17
28. Iams, J.D. (2010) Small for gestational age (SGA) and fetal growth restriction (FGR). *Am. J. Obstet. Gynecol.* 202, 513
29. Crispi, F. *et al.* (2018) Long-term cardiovascular consequences of fetal growth restriction: biology, clinical implications, and opportunities for prevention of adult disease. *Am. J. Obstet. Gynecol.* 218, S869–S879
30. Blencowe, H. *et al.* (2016) National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: a systematic analysis. *Lancet Glob. Health* 4, e98–e108
31. Smith, L.K. *et al.* (2018) Quantifying the burden of stillbirths before 28 weeks of completed gestational age in high-income countries: a population-based study of 19 European countries. *Lancet* 392, 1639–1646
32. Stillbirth Collaborative Research Network Writing Group (2011) Causes of death among stillbirths. *JAMA* 306, 2459–2468
33. Wall-Wieler, E. *et al.* (2019) Severe maternal morbidity among stillbirth and live birth deliveries in California. *Obstet. Gynecol.* 134, 310–317
34. Nonyane, B.A.S. *et al.* (2019) Pregnancy intervals after stillbirth, neonatal death and spontaneous abortion and the risk of an adverse outcome in the next pregnancy in rural Bangladesh. *BMC Pregnancy Childbirth* 19, 62
35. Ailes, E.C. *et al.* (2016) Using insurance claims data to identify and estimate critical periods in pregnancy: an application to antidepressants. *Birth Defects Res. A Clin. Mol. Teratol.* 106, 927–934
36. MacDonald, S.C. *et al.* (2019) Pregnancy outcomes in women with multiple sclerosis. *Am. J. Epidemiol.* 188, 57–66
37. Marić, I. *et al.* (2019) Data-driven queries between medications and spontaneous preterm birth among 2.5 million pregnancies. *Birth Defects Res.* 111, 1145–1153
38. Gao, C. *et al.* (2019) Deep learning predicts extreme preterm birth from electronic health records. *J. Biomed. Inform.* 100, 103334
39. Marić, I. *et al.* (2020) Early prediction of preeclampsia via machine learning. *Am. J. Obstet. Gynecol. MFM* 2, 100100
40. Qiu, H. *et al.* (2017) Electronic health record driven prediction for gestational diabetes mellitus in early pregnancy. *Sci. Rep.* 7, 16417
41. Boland, M.R. *et al.* (2017) Development of a machine learning algorithm to classify drugs of unknown fetal effect. *Sci. Rep.* 7, 12839
42. Challa, A.P. *et al.* (2020) EHRs could clarify drug safety in pregnant people. *Nat. Med.* 26, 820–821
43. Levy, O. *et al.* (2020) Early predictors of small-for-gestational-age neonates using non-invasive, low-cost, and readily available hematological markers. *Int. J. Gynaecol. Obstet.* 150, 340–345
44. Tanacan, A. *et al.* (2019) The importance of proteinuria in preeclampsia and its predictive role in maternal and neonatal outcomes. *Hypertens. Pregnancy* 38, 111–118
45. Kaduma, J. *et al.* (2019) Urinary tract infections and preeclampsia among pregnant women attending two hospitals in Mwanza City, Tanzania: a 1:2 matched case-control study. *Biomed. Res. Int.* 2019, 3937812
46. Burton, R.J. *et al.* (2019) Using artificial intelligence to reduce diagnostic workload without compromising detection of urinary tract infections. *BMC Med. Inform. Decis. Mak.* 19, 171
47. Coussons-Read, M.E. (2013) Effects of prenatal stress on pregnancy and human development: mechanisms and pathways. *Obstet. Med.* 6, 52–57
48. Weber, K.A. *et al.* (2020) Periconceptional stressors and social support and risk for adverse birth outcomes. *BMC Pregnancy Childbirth* 20, 487
49. Lilliecreutz, C. *et al.* (2016) Effect of maternal stress during pregnancy on the risk for preterm birth. *BMC Pregnancy Childbirth* 16, 5
50. Rejnö, G. *et al.* (2019) Maternal anxiety, depression and asthma and adverse pregnancy outcomes – a population based study. *Sci. Rep.* 9, 13101
51. Walsh, K. *et al.* (2019) Maternal prenatal stress phenotypes associate with fetal neurodevelopment and birth outcomes. *Proc. Natl. Acad. Sci. U. S. A.* 116, 23996–24005
52. Wu, G. *et al.* (2012) Biological mechanisms for nutritional regulation of maternal health and fetal development. *Paediatr. Perinat. Epidemiol.* 26, 4–26
53. AlSaif, S. *et al.* (2015) A short review of adipokines, smooth muscle and uterine contractility. *Life Sci.* 125, 2–8
54. Connor, K.L. *et al.* (2018) Maternal metabolic, immune, and microbial systems in late pregnancy vary with malnutrition in mice. *Biol. Reprod.* 98, 579–592
55. Carmichael, S.L. *et al.* (2018) Women's periconceptional diet and risk of biliary atresia in offspring. *Birth Defects Res.* 110, 994–1000
56. Weber, K.A. *et al.* (2018) Nutrient intake in women before conception and risks of anophthalmia and microphthalmia in their offspring. *Birth Defects Res.* 110, 863–870
57. Lu, M. *et al.* (2016) Association between vitamin D status and the risk of gestational diabetes mellitus: a meta-analysis. *Arch. Gynecol. Obstet.* 293, 959–966
58. Murthi, P. *et al.* (2016) Role of the placental vitamin D receptor in modulating feto-placental growth in fetal growth restriction and preeclampsia-affected pregnancies. *Front. Physiol.* 7, 43
59. Workalemahu, T. *et al.* (2018) Genetic and environmental influences on fetal growth vary during sensitive periods in pregnancy. *Sci. Rep.* 8, 7274

60. Mesner, O. *et al.* (2019) Using graph learning to understand adverse pregnancy outcomes and stress pathways. *PLoS One* 14, e0223319
61. Kim, M.K. *et al.* (2018) Socioeconomic status can affect pregnancy outcomes and complications, even with a universal healthcare system. *Int. J. Equity Health* 17, 2
62. Fasina, F. *et al.* (2020) Impact of mothers' socio-demographic factors and antenatal clinic attendance on neonatal mortality in Nigeria. *Cogent Soc. Sci.* 6, 1747328
63. Berger, K. *et al.* (2020) Prenatal phthalate, paraben, and phenol exposure and childhood allergic and respiratory outcomes: evaluating exposure to chemical mixtures. *Sci. Total Environ.* 725, 138418
64. Zheng, T. *et al.* (2016) Effects of environmental exposures on fetal and childhood growth trajectories. *Ann. Glob. Health* 82, 41–99
65. Patel, J. *et al.* (2020) Maternal occupational exposure to polycyclic aromatic hydrocarbons and the risk of isolated congenital heart defects among offspring. *Environ. Res.* 186, 109550
66. Choi, G. *et al.* (2019) Maternal exposure to outdoor air pollution and congenital limb deficiencies in the National Birth Defects Prevention Study. *Environ. Res.* 179, 108716
67. Stingone, J.A. *et al.* (2017) Maternal exposure to nitrogen dioxide, intake of methyl nutrients, and congenital heart defects in offspring. *Am. J. Epidemiol.* 186, 719–729
68. Zhang, G. *et al.* (2017) Genetic associations with gestational duration and spontaneous preterm birth. *N. Engl. J. Med.* 377, 1156–1167
69. Michita, R.T. *et al.* (2018) Genetic variants in preeclampsia: lessons from studies in Latin-American populations. *Front. Physiol.* 9, 1771
70. Huusko, J.M. *et al.* (2018) Whole exome sequencing reveals HSPA1L as a genetic risk factor for spontaneous preterm birth. *PLoS Genet.* 14, e1007394
71. Rappoport, N. *et al.* (2018) A genome-wide association study identifies only two ancestry specific variants associated with spontaneous preterm birth. *Sci. Rep.* 8, 226
72. Stevenson, D.K. *et al.* (2019) The contributions of genetics to premature birth. *Pediatr. Res.* 85, 416–417
73. Warrington, N.M. *et al.* (2019) Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.* 51, 804–814
74. Jaju Bhattad, G. *et al.* (2020) Histone deacetylase 1 and 2 drive differentiation and fusion of progenitor cells in human placental trophoblasts. *Cell Death Dis.* 11, 311
75. Kwak, Y.-T. *et al.* (2019) Human trophoblast differentiation is associated with profound gene regulatory and epigenetic changes. *Endocrinology* 160, 2189–2203
76. Wang, B. *et al.* (2019) Integrated RNA-seq and ChIP-seq analysis reveals a feed-forward loop regulating H3K9ac and key labor drivers in human placenta. *Placenta* 76, 40–50
77. Zhang, Y. *et al.* (2018) Comprehensive analysis of gene expression profiles and DNA methylome reveals Oas1, Ppie, Polr2g as pathogenic target genes of gestational diabetes mellitus. *Sci. Rep.* 8, 16244
78. Handelman, S.K. *et al.* (2019) The plasma metabolome of women in early pregnancy differs from that of non-pregnant women. *PLoS One* 14, e0224682
79. Goltsman, D.S.A. *et al.* (2018) Metagenomic analysis with strain-level resolution reveals fine-scale variation in the human pregnancy microbiome. *Genome Res.* 28, 1467–1480
80. Gomez-Lopez, N. *et al.* (2019) The cellular transcriptome in the maternal circulation during normal pregnancy: a longitudinal study. *Front. Immunol.* 10, 2863
81. Aghaeepour, N. *et al.* (2017) An immune clock of human pregnancy. *Sci. Immunol.* 2, eaan2946
82. Aghaeepour, N. *et al.* (2018) A proteomic clock of human pregnancy. *Am. J. Obstet. Gynecol.* 218, 347.e1–347.e14
83. Tarca, A.L. *et al.* (2020) Amniotic fluid cell-free transcriptome: a glimpse into fetal development and placental cellular dynamics during normal pregnancy. *BMC Med. Genet.* 13, 25
84. Tsang, J.C.H. *et al.* (2017) Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proc. Natl. Acad. Sci. U. S. A.* 114, E7786–E7795
85. Mei, C. *et al.* (2019) The unique microbiome and innate immunity during pregnancy. *Front. Immunol.* 10, 2886
86. Fettweis, J.M. *et al.* (2019) The vaginal microbiome and preterm birth. *Nat. Med.* 25, 1012–1021
87. Cobb, C.M. *et al.* (2017) The oral microbiome and adverse pregnancy outcomes. *Int. J. Womens Health* 9, 551–559
88. Yang, H. *et al.* (2020) Systematic analysis of gut microbiota in pregnant women and its correlations with individual heterogeneity. *NPJ Biofilms Microbiomes* 6, 32
89. Wainberg, M. *et al.* (2018) Deep learning in biomedicine. *Nat. Biotechnol.* 36, 829–838
90. Wang, D. (2001) Unsupervised learning: foundations of neural computation. *AI Mag.* 22, 101–102
91. Woolery, L.K. and Grzymala-Busse, J. (1994) Machine learning for an expert system to predict preterm birth risk. *J. Am. Med. Inform. Assoc.* 1, 439–446
92. Goodwin, L. and Maher, S. (2000) Data mining for preterm birth prediction. In *Proceedings of the 2000 ACM Symposium on Applied Computing – SAC '00*, pp. 46–51, ACM
93. Lu, Y. *et al.* (2019) Ensemble machine learning for estimating fetal weight at varying gestational age. *AAAI* 33, 9522–9527
94. Vovsha, I. *et al.* (2016) Using kernel methods and model selection for prediction of preterm birth. *Proc. Mach. Learn. Healthc.* 56, 55–72
95. Artzi, N.S. *et al.* (2020) Prediction of gestational diabetes based on nationwide electronic health records. *Nat. Med.* 26, 71–76
96. Malacova, E. *et al.* (2020) Stillbirth risk prediction using machine learning for a large cohort of births from Western Australia, 1980–2015. *Sci. Rep.* 10, 5354
97. Ye, Y. *et al.* (2020) Comparison of machine learning methods and conventional logistic regressions for predicting gestational diabetes using routine clinical data: a retrospective cohort study. *J. Diabetes Res.* 2020, 4168340
98. Kuhle, S. *et al.* (2018) Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. *BMC Pregnancy Childbirth* 18, 333
99. Koivu, A. and Sairanen, M. (2020) Predicting risk of stillbirth and preterm pregnancies with machine learning. *Health Inf. Sci. Syst.* 8, 14
100. Catley, C. *et al.* (2006) Predicting high-risk preterm birth using artificial neural networks. *IEEE Trans. Inform. Technol. Biomed.* 10, 540–549
101. Sufriyana, H. *et al.* (2020) Prediction of preeclampsia and intrauterine growth restriction: development of machine learning models on a prospective cohort. *JMIR Med. Inform.* 8, e15411
102. Wright, D. *et al.* (2019) Predictive performance of the competing risk model in screening for preeclampsia. *Am. J. Obstet. Gynecol.* 220, 199.e1–199.e13
103. Akbulut, A. *et al.* (2018) Fetal health status prediction based on maternal clinical history using machine learning techniques. *Comput. Methods Prog. Biomed.* 163, 87–100
104. Rawashdeh, H. *et al.* (2020) Intelligent system based on data mining techniques for prediction of preterm birth for women with cervical cerclage. *Comput. Biol. Chem.* 85, 107233
105. Chen, H.-Y. *et al.* (2011) Exploring the risk factors of preterm birth using data mining. *Expert Syst. Appl.* 38, 5384–5387
106. Zhang, W. *et al.* (2020) Machine learning models for the prediction of postpartum depression: application and comparison based on a cohort study. *JMIR Med. Inform.* 8, e15516
107. Saeys, Y. *et al.* (2007) A review of feature selection techniques in bioinformatics. *Bioinformatics* 23, 2507–2517
108. Cunningham, J.P. and Ghahramani, Z. (2015) Linear dimensionality reduction: survey, insights, and generalizations. *J. Mach. Learn. Res.* 16, 2859–2900
109. Hastie, T. *et al.* (2015) Statistical inference. In *Statistical Learning with Sparsity: The Lasso and Generalizations*, pp. 155–182, Chapman & Hall/CRC
110. Culos, A. *et al.* (2020) Integration of mechanistic immunological knowledge into a machine learning pipeline improves predictions. *Nat. Mach. Intell.* 2, 619–628
111. Hong, S. *et al.* (2019) Longitudinal profiling of human blood transcriptome in healthy and lupus pregnancy. *J. Exp. Med.* 216, 1154–1169

112. Jelliffe-Pawlowski, L.L. *et al.* (2018) Prediction of preterm birth with and without preeclampsia using mid-pregnancy immune and growth-related molecular factors and maternal characteristics. *J. Perinatol.* 38, 963–972
113. Guo, Z. *et al.* (2020) Whole-genome promoter profiling of plasma DNA exhibits diagnostic value for placenta-origin pregnancy complications. *Adv. Sci. (Weinh.)* 7, 1901819
114. Ghaemi, M.S. *et al.* (2019) Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy. *Bioinformatics* 35, 95–103
115. Wolpert, D.H. (1992) Stacked generalization. *Neural Netw.* 5, 241–259
116. Li, E. *et al.* (2019) Identification of diagnostic biomarkers in patients with gestational diabetes mellitus based on transcriptome gene expression and methylation correlation analysis. *Reprod. Biol. Endocrinol.* 17, 112
117. Lee, B. *et al.* (2019) Heterogeneous multi-layered network model for omics data integration and analysis. *Front. Genet.* 10, 1381
118. Koh, H.W.L. *et al.* (2019) iOmicsPASS: network-based integration of multiomics data for predictive subnetwork discovery. *NPJ Syst. Biol. Appl.* 5, 22
119. Yan, J. *et al.* (2018) Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. *Brief. Bioinform.* 19, 1370–1381
120. Wu, C. *et al.* (2019) A selective review of multi-level omics data integration using variable selection. *High Throughput* 8, 4
121. Zhu, L. *et al.* (2019) Bayesian indicator variable selection to incorporate hierarchical overlapping group structure in multi-omics applications. *Ann. Appl. Stat.* 13, 2611–2636
122. Fang, Z. *et al.* (2018) Bayesian integrative model for multi-omics data with missingness. *Bioinformatics* 34, 3801–3808
123. Ronen, J. *et al.* (2019) Evaluation of colorectal cancer subtypes and cell lines using deep learning. *Life Sci. Alliance* 2, e201900517
124. Zhang, L. *et al.* (2018) Deep learning-based multi-omics data integration reveals two prognostic subtypes in high-risk neuroblastoma. *Front. Genet.* 9, 477
125. Chung, N.C. *et al.* (2019) Unsupervised classification of multi-omics data during cardiac remodeling using deep learning. *Methods* 166, 66–73
126. Nguyen, N.D. and Wang, D. (2020) Multiview learning for understanding functional multiomics. *PLoS Comput. Biol.* 16, e1007677
127. Stanley, N. *et al.* (2020) VoPo leverages cellular heterogeneity for predictive modeling of single-cell data. *Nat. Commun.* 11, 3738
128. Butte, A. (2020) A robust and interpretable, end-to-end deep learning model for cytometry data. *J. Immunol.* 204, 159.8
129. Marzbanrad, F. *et al.* (2018) Cardiotocography and beyond: a review of one-dimensional Doppler ultrasound application in fetal monitoring. *Physiol. Meas.* 39, 08TR01
130. Signorini, M.G. *et al.* (2020) Integrating machine learning techniques and physiology based heart rate features for antepartum fetal monitoring. *Comput. Methods Prog. Biomed.* 185, 105015
131. Ismail Fawaz, H. *et al.* (2019) Deep learning for time series classification: a review. *Data Min. Knowl. Discov.* 33, 917–963
132. Bagnall, A. *et al.* (2017) The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min. Knowl. Discov.* 31, 606–660
133. Harutyunyan, H. *et al.* (2019) Multitask learning and benchmarking with clinical time series data. *Sci. Data* 6, 96
134. Zhao, Z. *et al.* (2019) DeepFHR: intelligent prediction of fetal acidemia using fetal heart rate signals based on convolutional neural network. *BMC Med. Inform. Decis. Mak.* 19, 286
135. Facco, F.L. *et al.* (2018) Association of adverse pregnancy outcomes with self-reported measures of sleep duration and timing in women who are nulliparous. *J. Clin. Sleep Med.* 14, 2047–2056
136. Souza, R.T. *et al.* (2019) Identification of earlier predictors of pregnancy complications through wearable technologies in a Brazilian multicentre cohort: Maternal Actigraphy Exploratory Study I (MAES-I) study protocol. *BMJ Open* 9, e023101
137. Carroll, J.E. *et al.* (2020) Sleep disturbances and inflammatory gene expression among pregnant women: differential responses by race. *Brain Behav. Immun.* 88, 654–660
138. Khademi, A. *et al.* (2019) Personalized sleep parameters estimation from actigraphy: a machine learning approach. *Nat. Sci. Sleep* 11, 387–399
139. Tsai, S.-Y. *et al.* (2012) Daily sleep and fatigue characteristics in nulliparous women during the third trimester of pregnancy. *Sleep* 35, 257–262
140. Aggarwal, J.K. and Ryoo, M.S. (2011) Human activity analysis. *ACM Comput. Surv.* 43, 1–43
141. Nweke, H.F. *et al.* (2018) Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. *Expert Syst. Appl.* 105, 233–261
142. Wang, J. *et al.* (2018) Deep learning for sensor-based activity recognition: a survey. *Pattern Recogn. Lett.* 119, 3–11
143. Ziaeefard, M. and Bergevin, R. (2015) Semantic human activity recognition: a literature review. *Pattern Recogn.* 48, 2329–2345
144. Shen, D. *et al.* (2017) Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 19, 221–248
145. Litjens, G. *et al.* (2017) A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88
146. Liu, S. *et al.* (2019) Deep learning in medical ultrasound analysis: a review. *Engineering* 5, 261–275
147. Baharlou, H. *et al.* (2019) Mass cytometry imaging for the study of human diseases – applications and data analysis strategies. *Front. Immunol.* 10, 2657
148. Whitworth, M. *et al.* (2015) Ultrasound for fetal assessment in early pregnancy. *Cochrane Database Syst. Rev.* 2015, CD007058
149. Pagani, G. *et al.* (2014) Association between reduced fetal movements at term and first trimester markers of impaired placental development. *Placenta* 35, 606–610
150. Sharp, A. *et al.* (2019) A prediction model for short-term neonatal outcomes in severe early-onset fetal growth restriction. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 241, 109–118
151. Rawat, V. *et al.* (2018) Automated techniques for the interpretation of fetal abnormalities: a review. *Appl. Bionics Biomech.* 2018, 6452050
152. Attallah, O. *et al.* (2018) Detecting and classifying fetal brain abnormalities using machine learning techniques. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1371–1376, IEEE
153. Arnaout, R. *et al.* (2018) Deep-learning models improve on community-level diagnosis for common congenital heart disease lesions. *arXiv* Published online September 19, 2018. https://arxiv.org/abs/1809.06993
154. Shickel, B. *et al.* (2018) Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J. Biomed. Health Inform.* 22, 1589–1604
155. Koleck, T.A. *et al.* (2019) Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J. Am. Med. Inform. Assoc.* 26, 364–379
156. Gehrmann, S. *et al.* (2018) Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. *PLoS One* 13, e0192360
157. Westergaard, D. *et al.* (2018) A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. *PLoS Comput. Biol.* 14, e1005962
158. Gonzalez, G.H. *et al.* (2016) Recent advances and emerging applications in text and data mining for biomedical discovery. *Brief. Bioinform.* 17, 33–42
159. Zweigenbaum, P. *et al.* (2007) Frontiers of biomedical text mining: current progress. *Brief. Bioinform.* 8, 358–375
160. Wang, X. *et al.* (2019) Cross-type biomedical named entity recognition with deep multi-task learning. *Bioinformatics* 35, 1745–1752
161. Craven, M. and Kumlien, J. (1999) Constructing biological knowledge bases by extracting information from text sources. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1999, 77–86
162. Zunic, A. *et al.* (2020) Sentiment analysis in health and well-being: systematic review. *JMIR Med. Inform.* 8, e16023
163. Zhang, L. *et al.* (2018) Deep learning for sentiment analysis: a survey. *WIREs Data Min. Knowl. Discov.* 8, e1253
164. McCoy, T.H. *et al.* (2015) Sentiment measured in hospital discharge notes is associated with readmission and mortality

risk: an electronic health record study. *PLoS One* 10, e0136341
165. Li, Y. *et al.* (2019) A survey of multi-view representation learning. *IEEE Trans. Knowl. Data Eng.* 31, 1863–1883
166. Baltrusaitis, T. *et al.* (2019) Multimodal machine learning: a survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 423–443
167. Ngiam, J. *et al.* (2011) Multimodal deep learning. In *28th International Conference on Machine Learning*, pp. 689–696, ICML
168. Zhao, J. *et al.* (2017) Multi-view learning overview: recent progress and new challenges. *Inf. Fusion* 38, 43–54
169. Li, Y. *et al.* (2018) A review on machine learning principles for multi-view biological data integration. *Brief. Bioinform.* 19, 325–340
170. Wang, W. *et al.* (2015) *On deep multi-view representation learning*. Proceedings of the 32nd International Conference on Machine Learning, PMLR 37 pp. 1083–1092
171. Guo, W. *et al.* (2019) Deep multimodal representation learning: a survey. *IEEE Access* 7, 63373–63394
172. Ruder, S. (2017) An overview of multi-task learning in deep neural networks. *arXiv* Published online June 15, 2017. https://arxiv.org/abs/1706.05098
173. Caruana, R. (1997) Multitask learning. *Mach. Learn.* 28, 41–75
174. Zheng, L. *et al.* (2019) Deep multimodality model for multi-task multi-view learning. In *Proceedings of the 2019 SIAM International Conference on Data Mining* (Berger-Wolf, T. and Chawla, N., eds), pp. 10–18, Society for Industrial and Applied Mathematics
175. Leclercq, M. *et al.* (2019) Large-scale automatic feature selection for biomarker discovery in high-dimensional omics data. *Front. Genet.* 10, 452
176. Cox, B. *et al.* (2015) Placental transcriptome in development and pathology: expression, function, and methods of analysis. *Am. J. Obstet. Gynecol.* 213, S138–S151
177. Ibarra, A. *et al.* (2020) Non-invasive characterization of human bone marrow stimulation and reconstitution by cell-free messenger RNA sequencing. *Nat. Commun.* 11, 400
178. Johnson, C.H. *et al.* (2016) Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* 17, 451–459
179. Timp, W. and Timp, G. (2020) Beyond mass spectrometry, the next step in proteomics. *Sci. Adv.* 6, eaax8978
180. Knight, R. *et al.* (2017) The microbiome and human biology. *Annu. Rev. Genomics Hum. Genet.* 18, 65–86
181. Zou, H. and Hastie, T. (2005) Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat Methodol.* 67, 301–320
182. Kotsiantis, S.B. (2013) Decision trees: a recent overview. *Artif. Intell. Rev.* 39, 261–283
183. Breiman, L. (2001) Random forests. *Mach. Learn.* 45, 5–32
184. Chen, T. and Guestrin, C. (2016) XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD '16*, pp. 785–794, SIGKDD
185. Hsu, C.-W. and Lin, C.-J. (2002) A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* 13, 415–425
186. Altman, N.S. (1992) An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* 46, 175–185
187. Friedman, N. *et al.* (1997) Bayesian network classifiers. *Mach. Learn.* 29, 131–163
188. Miotto, R. *et al.* (2018) Deep learning for healthcare: review, opportunities and challenges. *Brief. Bioinform.* 19, 1236–1246
189. Eraslan, G. *et al.* (2019) Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.* 20, 389–403
190. Cheerla, A. and Gevaert, O. (2019) Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics* 35, i446–i454