Artificial Intelligence In Medicine 181 (2026) 103510

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Research paper

PregBase: A comprehensive knowledge base for clinically relevant knowledge representation and biomarker prediction in pregnancy

Aashish Bhandari [a], Deval Mehta [a,b], Karin Verspoor [a], Damiano Spina [a], Feng Xia [a], Sonika Tyagi [a,c,*]

[a] Data Science and AI Division, School of Computing Technologies, RMIT University, Melbourne 3000, VIC, Australia
[b] Department of Data Science & AI, Faculty of IT, Monash University, Melbourne 3800, VIC, Australia
[c] The Alfred Hospital and School of Translational Medicine, Monash University, Prahran 3004, VIC, Australia

ARTICLE INFO

Dataset link: https://gitlab.com/tyagilab/pregbase, https://pregknowledgebase.com

Keywords:
Knowledge graph
Pregnancy complications
Large language model
Biomarker prediction
Link prediction
Maternal health
Information extraction
Graph learning

ABSTRACT

Pregnancy complications are a leading cause of maternal and neonatal mortality worldwide. Understanding their underlying mechanisms is hindered by dispersed evidence across thousands of studies and complex biological interactions. We present *PregBase*, a comprehensive knowledge base for pregnancy research comprising automated extraction, validation, and exploration. PregBase was constructed by utilising large language models (LLMs) to extract relationships from literature; 13 LLMs were benchmarked across seven prompting strategies, with ensemble shuffle prompting outperforming single-strategy alternatives. A three-tier validation pipeline combining ontology mapping, graph neural networks, and statistical analysis produced *PregKG*, a knowledge graph containing 64,087 associations between 13,303 biomedical entities spanning 155 semantic types and 50 vocabularies across 8 relationship types from 8420 articles. Link prediction validated PregBase's inference capability beyond extracted knowledge, recovering established clinical interventions, reconstructing canonical hormonal pathways across maternal–placental–fetal compartments, and identifying novel biomarker candidates for preterm birth, gestational diabetes, and preeclampsia, supported by genetic and expression databases. An interactive web interface (https://pregknowledgebase.com) has been created to enable further exploration via conversational queries. This work provides a scalable foundation for systematic discovery in maternal health research.

1. Introduction

Human pregnancy is associated with a high rate of abnormal outcomes. Adverse pregnancy outcomes such as preterm birth, gestational diabetes, and preeclampsia remain major causes of maternal and neonatal complications worldwide [1–3]. The effects of these outcomes extend beyond the immediate postpartum period, having a long-term effect on both the mother and child and placing a huge burden on healthcare systems and families due to prolonged hospital stays, intensive care, and long-term treatments [4]. According to WHO, approximately 810 women die daily from pregnancy-related complications, with the majority occurring in low- and middle-income countries [5]. Despite recent advances in medical technology reducing maternal mortality rates [6,7], predicting and preventing pregnancy complications continues to challenge clinicians and researchers [8]. Most obstetric complications are preventable with timely assessment and appropriate intervention [9,10], but there is a struggle to identify at-risk individuals early enough. Thus, it is essential to understand the biological mechanisms driving these complications to enable the development of effective prevention strategies.

Biomedical literature is expanding rapidly, with thousands of studies published annually on pregnancy outcomes and maternal health research. The volume of new publications daily makes manual synthesis impractical. Thus, enormous quantities of research data generated remain underutilised [11]. Similarly, critical knowledge is fragmented and dispersed, not only in unstructured text across scientific literature but also across clinical trial registries, clinical guidelines, and public health documentation [11,12]. There is a need to compile structured pregnancy research knowledge as medical understanding grows. Understanding pregnancy complications requires connecting information scattered throughout numerous studies; however, there is no computational infrastructure to link these fragments systematically. Furthermore, multiple medical conditions frequently co-occur during pregnancy, but when treating these complex cases, clinicians must manually piece together information from disparate sources to

* Corresponding author at: Data Science and AI Division, School of Computing Technologies, RMIT University, Melbourne 3000, VIC, Australia.
E-mail address: sonika.tyagi@rmit.edu.au (S. Tyagi).

https://doi.org/10.1016/j.artmed.2026.103510
Received 1 February 2026; Received in revised form 19 July 2026; Accepted 12 August 2026
Available online 19 August 2026

understand potential risks. For example, a woman with diabetes, hypertension, and a history of miscarriage requires an integrated assessment of how these factors interact, but existing systems treat each condition separately [13,14]. Without structured integration of pregnancy knowledge, healthcare researchers cannot systematically trace biological mechanisms or synthesise evidence across the scale of existing literature.

Knowledge extraction from biomedical literature has shifted from rule-based systems to neural approaches. Early methods used predefined patterns and medical vocabularies [15,16] but struggled with the diverse ways researchers describe findings. Transformer models improved this with domain-trained models like BioBERT [17] and SciBERT [18] supporting more flexible recognition of medical concepts and relationships. Joint extraction approaches simultaneously identify entities and relations [18,19], whilst ensemble methods combine different pretrained models for entity recognition followed by separate relation extraction [20]. Other approaches integrated structured parsing with pretrained models [21,22], though they required complex multi-stage pipelines and have primarily focused on abstracts rather than full-text articles. Decoder-based large language models (LLMs) have since enabled end-to-end extraction by identifying complete triples through prompting alone, without task-specific training [23], with researchers generally relying on zero-shot or few-shot strategies [24–26]. As prompt design significantly affects extraction performance [27], LLMs have enabled end-to-end extraction through prompting strategies, scaling to millions of relationships in materials science [28] and mental health [23]. Despite these advances in general biomedical domains, pregnancy complications remain underexplored, and none of these approaches have been systematically evaluated for pregnancy-specific literature at scale.

Knowledge graphs (KGs) have emerged as critical tools for organising biomedical relationships as structured, interconnected networks by integrating large volumes of heterogeneous data, thereby enabling efficient information retrieval and automated knowledge discovery [11]. Large-scale resources have been developed, including SPOKE, which integrates 41 structured biomedical databases [29], MedKG, a drug-focused KG with link prediction support [22], and CHDbase, which combines gene-phenotype relationships for congenital heart disease [30]. Additional resources like UMLS, SemMedDB, and PrimeKG have proven to be valuable for drug repurposing, disease pathway analysis, and biomarker discovery [31–33]. They enable automated literature mining, integration of multi-omics data, and identification of novel disease-gene associations. However, existing computational approaches face key limitations. Machine learning models can predict pregnancy outcomes with reasonable accuracy [34,35], but they function as black boxes that forecast risk without explaining the underlying biological mechanisms. Systematic reviews highlight that computational systems can predict outcomes but cannot reveal the understanding and reasoning behind the complications or interactions of biological pathways [34]. Early studies identified risk factors for conditions such as preterm birth [36], but translating statistical associations into biological insights remains a challenge. Moreover, while existing KG systems provide valuable resources, they largely rely on structured databases and manual curation, limiting their ability to incorporate emerging knowledge from newly published literature. Existing KG systems also cannot be directly adapted to the domain of pregnancy due to the following limitations. First, they provide sparse coverage of pregnancy-specific concepts such as maternal–fetal interactions, gestational-age-dependent changes, and pregnancy-specific interventions. Moreover, most KGs represent information as simple triples like *Drug, treats, Disease* that cannot capture complex, nested relationships where subjects and objects are themselves complete facts [37,38]. Second, they do not account for contextualised relationships as each woman's medical conditions, lifestyle, and genetics influence outcomes differently; for example, pre-existing diabetes increases birth defect risk threefold [39], and effects vary by trimester and parity. Third, existing biomedical KGs rely primarily on structured databases and manual curation [22,29], limiting their ability to capture emerging knowledge from rapidly expanding literature.

To address these limitations, we present the pregnancy knowledge base (**PregBase**), the first knowledge base built specifically for pregnancy research. While resources such as SPOKE [29] and PrimeKG [40] include some pregnancy conditions as general disease nodes, neither is designed for pregnancy-specific concepts such as maternal-fetal interactions or gestational-age-dependent changes. PregBase is built upon the pregnancy knowledge graph (PregKG), which represents the underlying graph structure of entities and relationships extracted from literature. PregBase extends beyond a traditional KG by integrating automated extraction, evaluation and reasoning capabilities, complemented with multiple interaction tools. We benchmark 13 LLMs across seven prompting strategies, going beyond the simple zero-shot or few-shot approaches used in previous work [20,28,41]. This provides a comprehensive evaluation of LLM-based biomedical triple extraction that contributes transferable insights for KG construction beyond model selection. Specifically, to broaden the coverage of pregnancy-specific features, the extracted triples undergo multi-stage validation through three steps: (1) Ontology-guided mapping using UMLS [42] for entity normalisation; (2) Scoring based on a trained graph neural network (GNN) [43] on verified relationships; and (3) Statistical validation via pointwise mutual information (PMI). Furthermore, to capture the contextualised nature of pregnancy biology, we construct nested triple structures that preserve how multiple factors interact within specific conditions. More importantly, PregBase makes this knowledge accessible through an interactive web interface with graph visualisation, entity searches, and semantic chain exploration, alongside a chat assistant for intuitive querying without programming expertise. The system is designed for biomedical researchers and clinician-scientists seeking to trace biological mechanisms and prioritise candidates for experimental follow-up. We demonstrate this through link prediction, identifying novel biomarker candidates for preterm birth, gestational diabetes, and preeclampsia as computationally derived predictions requiring experimental validation. PregBase is intended as a research tool for systematic knowledge exploration in maternal health, complementing rather than replacing clinical judgment. Overall, PregBase has three components: PregKG as the underlying data, link prediction for biomarker identification, and a web interface for research exploration. The system targets hypothesis generation in maternal health research, rather than clinical decision support. Table 1 compares our developed PregBase with representative biomedical knowledge bases, highlighting its novel capabilities, including LLM-based triple extraction, ontology-guided graph construction and statistical-based validation, nested triple construction, and a web-based interface with a natural language chat assistant.

2. Methods

Our automated and modular pipeline workflow comprises five stages and is shown in Fig. 1. Specifically, we developed this pipeline to extract, validate, and structure biomedical knowledge from unstructured text, regardless of its source. The system integrates literature ingestion, LLM-based triple extraction, multi-tier testing, and web deployment. More importantly, our architecture is designed to be domain-agnostic and readily adaptable.

2.1. Automated literature ingestion and filtering

Pregnancy-related articles were collected from PubMed using targeted keywords such as *"preterm birth", "gestational diabetes" and "pregnancy AND foetal development"*. The search was restricted to English-language human studies published from 1960 to 2025, though some animal studies may be present given the automated retrieval. *Biopython*

Table 1
Comparison of existing biomedical knowledge bases with our PregBase on different aspects.

KG	Data Source	Scope	Extraction	Validation Method	Reasoning Capabilities	Interface
CHDbase (2023) [30]	Curated gene-phenotype associations	Congenital heart disease	Manual	Manual review	None	Web interface with search
SPOKE (2023) [29]	Structured biomedical databases	General biomedical	Structured parsing	Ontology mapping + rule-based filtering	Basic graph traversal	Web interface with search and neighbour exploration
PrimeKG (2023) [40]	Structured biomedical databases + curated text	General biomedical	Structured parsing	Ontology mapping + embedding-based grouping	Network analysis	Web interface with search (Neo4j code)
LMKG (2024) [21]	Structured biomedical databases + unstructured text	General biomedical	pretrained NER/RE models + LLM-assisted alignment	Entity alignment + manual review	KG embeddings + clinical decision support	Interactive web interface with search and network visualisation
MedKG (2025) [22]	Structured biomedical databases + unstructured text	General biomedical (drug-focused)	Structured parsing + pretrained NER/RE models + LLM-assisted classification	Ontology mapping + rule-based filtering	Link prediction	Interactive web interface with search and network visualisation
PregBase (Ours)	Unstructured pregnancy-related text (abstract, full-text)	Pregnancy complications	Prompt-based LLM extraction with modifier support	Hybrid (ontology mapping, graph scoring and statistical validation)	Link prediction, nested relationships & modifier-aware reasoning	Interactive web interface with search, network visualisation, and natural language assistant

*__NER__: Named Entity Recognition; __RE__: Relation Extraction. Systems differ in scope and design goals. The comparison reflects architectural differences rather than performance.

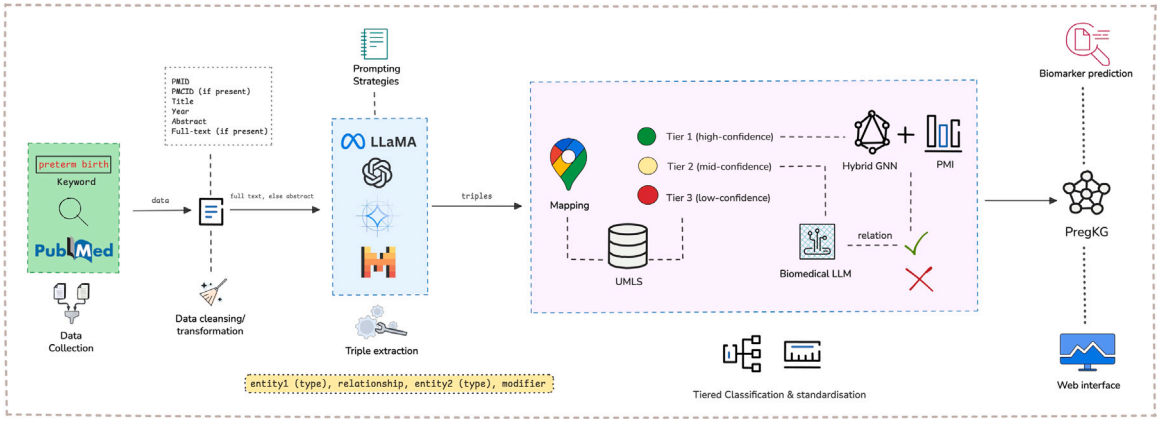


Fig. 1. Overview of the PregBase pipeline. The workflow comprises five main stages: (1) data collection from PubMed using pregnancy-related keywords, (2) preprocessing and text formatting, (3) LLM-based triple extraction with benchmarked prompting strategies, (4) triple standardisation and tiered classification through UMLS mapping, biomedical LLM-based plausibility adjudication (OpenBioLLM-8B), hybrid graph neural network scoring, and PMI-based statistical validation, and (5) downstream applications including an interactive web interface and biomarker prediction.

[44] extracted metadata containing title, PMID, PMCID, year of publication, and abstract. The full text was added only when a PubMed Central (PMC) version was available, as not all records provide this. Tables, figures and supplementary materials were removed to focus on the main content of the article. The corpus was stored as structured JSON (8420 articles).

2.2. Triple extraction with LLMs

Articles were processed to extract biomedical relationships as triples: $(\text{entity}_1\ (\text{type}),\ \text{relation},\ \text{entity}_2\ (\text{type}),\ \text{modifier})$. Relation types were adopted from the BioRED relation schema [20,45] and constrained in the extraction prompt to eight types: *Association, Positive_Correlation, Negative_Correlation, Bind, Comparison, Conversion, Cotreatment,* and *Drug_Interaction*. These types map broadly to Biolink Model predicates [46], though BioRED was used as the source schema given its derivation from biomedical literature rather than top-down ontological design. Relation labels were validated post-extraction against these eight permitted types. Triples with out-of-schema labels were excluded from the final graph. Entity types were initially restricted to 19 UMLS semantic categories and subsequently validated through UMLS ontology mapping during triple classification. Modifiers capture optional contextual information about a relationship, such as negation, severity, quantification, or temporal context, and are stored as edge attributes alongside each triple. We benchmarked several open-source LLMs in seven prompting strategies on the BioRED [45] and BioREL [47] datasets. Full prompt templates, strategy definitions, and model version strings are provided in Supplementary Section 1. All models were run at a temperature of 0.4 rather than 0. At temperature

= 0, shuffled prompt variants produce identical outputs and the ensemble effect is lost. Each model–prompt combination was assessed for its ability to correctly identify entity spans (named-entity recognition (NER)), assign appropriate semantic types, and predict valid relations between entities. The optimal configuration (highest macro-averaged F1 across all tasks) was selected for PregKG construction.

2.3. Tiered classification framework

The extracted triples underwent a three-stage assessment. First, ontology-guided mapping aligned entities with UMLS concepts to assign Concept Unique Identifiers (CUIs), semantic types, and vocabulary sources, classifying triples into confidence-based tiers. Validation statistics, including entity and relation matching accuracy, triple-level accuracy, and MRR, were collected to compare biomedical embedding models. Second, graph-based scoring assessed structural consistency with known biomedical relationships. Third, statistical validation evaluated entity pair co-occurrence in the literature. Together, these stages prioritised triples for graph construction.

Ontology-guided Mapping: Each triple was validated on three criteria: entity grounding, semantic type consistency, and relationship verification. Entities were embedded using SapPMBERT [48], a biomedical language model shown to perform well on UMLS concept alignment. The resulting embeddings were compared with UMLS concept embeddings using cosine similarity to assign CUIs. This process also returned the semantic type of each entity, along with the vocabulary and source information from UMLS. Relationship validation was performed using the UMLS *MRREL* table, which specifies the known relations between mapped CUIs. Semantic type validation used the *MRSTY* table to ensure predicted entity type consistency with UMLS categories. This process ensured each triple remained grounded in standard biomedical terminology with reliable metadata.

The validated triples were classified into three confidence-based tiers. Tier 1 represented high-confidence triples, where both entities mapped to UMLS CUIs with correct semantic types and the relation was present in UMLS MRREL. Tier 2 included mid-confidence triples, where both entities mapped to CUIs with correct types but no corresponding relation existed in MRREL; these were further reviewed by OpenBioLLM-8B [49] to judge biological plausibility, using deterministic inference (temperature 0) with the prompt provided in Supplementary Section 1. OpenBioLLM-8B is openly available on HuggingFace with no proprietary access required. The model was selected through benchmarking against six candidate models on a held-out subset, balancing Tier 2 and Tier 3 agreement rates (Supplementary Table S2). Tier 3 covered low-confidence triples, where neither mapping nor LLM judgment confirmed the prediction; these were excluded from further analysis. This tiered design balanced standardisation through UMLS with semantic flexibility through LLMs. The resulting triples formed the foundation of the KG.

Tier 2 Refinement Through Hybrid Scoring: The mid-confidence Tier 2 triples typically outnumbered Tier 1. To refine them, we trained a hybrid GNN combining graph attention network (GAT), relational graph convolutional network (R-GCN) and graph convolutional network (GCN) layers. The model was trained on the Tier 1 subset, which served as a high-confidence reference graph. Each entity was represented using type-aware node features, with relation-specific message-passing encoding of structural context. The trained model scored Tier 2 triples, estimating their plausibility based on graph structure and learned embeddings. These scores provided one axis of validation.

To further assess Tier 2 triples, we computed PMI scores between entity pairs using the pregnancy literature corpus. Entity pairs that co-occur frequently in the literature are more likely to share genuine biological relationships, consistent with PMI-based biomedical literature mining [50]. Co-occurrence was measured using a sliding sentence window of size 2. PMI was computed as:

$$\mathrm{PMI}(e_1, e_2) = \log \frac{P(e_1, e_2)}{P(e_1) \cdot P(e_2)} \tag{1}$$

where $P(e_1, e_2)$ is the joint co-occurrence probability and $P(e_1)$, $P(e_2)$ are marginal probabilities across all sentence windows. Scores were normalised to [0,1] using min–max scaling to align with graph-based plausibility scores.

The final confidence score combined the graph and PMI components:

$$\mathrm{Score} = \alpha \cdot \mathrm{Graph} + (1 - \alpha) \cdot \mathrm{PMI} \tag{2}$$

The graph score receives the higher weight because the GNN is trained on pregnancy-specific Tier 1 triples and captures domain-specific relational structure, whereas PMI reflects general literature co-occurrence patterns. Rather than applying a fixed threshold, we adopted a top-k percentile selection strategy, retaining only the most confident triples based on the combined score distribution. The threshold is selected to balance precision against coverage. These selected Tier 2 triples were combined with Tier 1 to create the validated set for downstream tasks, including nested triple construction and link prediction.

2.4. Capturing deeper biomedical relationships

From the validated set, we construct *nested triples*, where the object of one relation also serves as the subject of another. This creates *semantic chains* that capture multi-hop connections between biomedical entities and allow reasoning beyond single-step links. The general structure is `(Outer Subject, Outer Relation, (Inner Subject, Inner Relation, Inner Object))`: for example, *(Drug A, inhibits, (Protein B, activates, Gene C))*. These nested triples are stored alongside flat triples to enable reasoning beyond single-step relationships. Chains are identified when the object of one triple matches the subject of another, triggering a combination into a nested structure. Each nested triple is stored as a flat record with explicit outer and inner subject, relation, object, semantic type, modifier, and vocabulary fields. Downstream link prediction models operate on flat triples only, as standard embedding methods do not natively support nested structures. These structures are accessible through the semantic chain exploration feature of the web interface, described in Section 3.6. The full graph schema, from entity attributes to nested triple structure, is illustrated in Fig. 2.

2.5. Graph reasoning with link prediction

The KG's reasoning capability was evaluated through link prediction experiments, which assess whether models can infer missing relationships between entities based on existing knowledge. Articles were sorted chronologically by publication date, with the earliest 80% of triples allocated to training, 10% to validation, and the most recent 10% to testing. This temporal split simulated realistic knowledge-discovery scenarios in which models predict relationships in recent literature based on historical evidence [51].

Three categories of models were compared: embedding-based methods (TransE [52], DistMult [53], and ComplEx [54]), GNNs (CompGCN [55], R-GCN [56]), and hybrid architectures combining GAT, R-GCN, and residual connections [22]. Three hybrid variants were evaluated. V1 uses a standard GAT layer without relation awareness, followed by R-GCN and a residual GCN layer, trained with binary cross-entropy loss. V2 replaces the standard GAT with a relation-aware GATv2 layer and adds a contrastive loss term alongside binary cross-entropy. V2-Edge retains the relation-aware GATv2 but replaces the combined loss with an edge-side relation bias injection, where relation embeddings are added to source node features before message passing, and trains with binary cross-entropy only. In all variants, entity and relation embeddings have dimension 200, with 20 basis vectors for R-GCN relation decomposition, one hidden layer, and dropout of 0.2. Initial experiments on a subset of PregKG and scalability validation on PrimeKG [40], a large external KG, identified optimal configurations before application to the complete dataset.

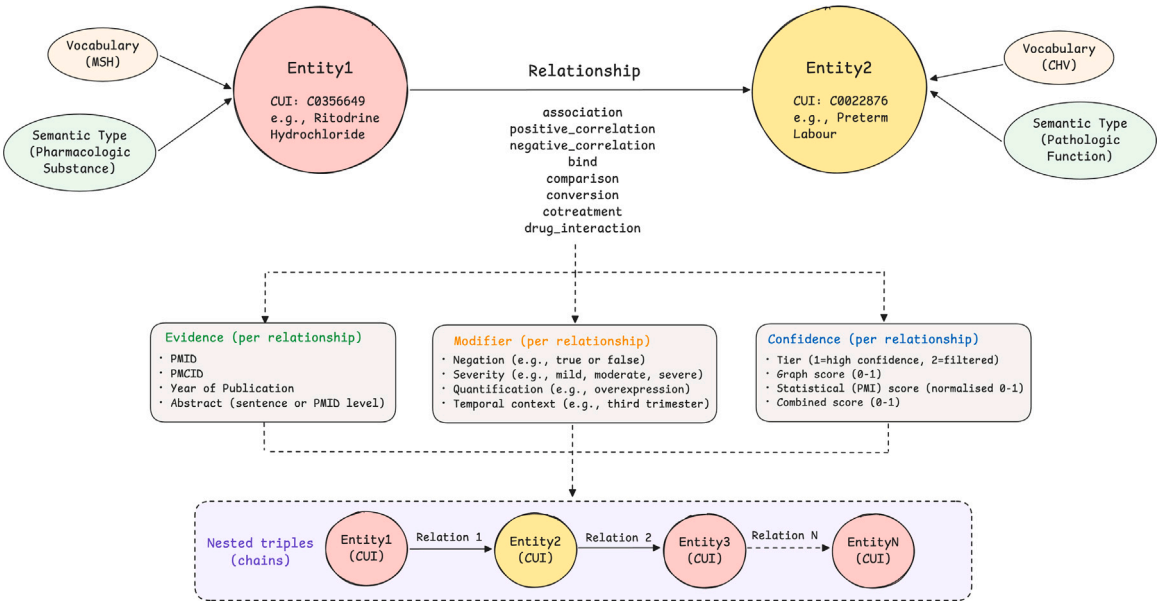


Fig. 2. PregKG schema. Entity nodes carry a CUI, UMLS semantic type, and vocabulary source. Edges represent one of eight relation types and carry three attribute groups: evidence, modifier, and confidence. The lower panel shows the nested triple structure, where entities are chained via sequential relations to encode multi-hop biological pathways.

Models were trained using binary cross-entropy loss with negative sampling at a 1:20 positive-to-negative ratio, Adam optimiser (learning rate 0.001), batch size 128, and 20 epochs. Negative triples were generated by uniform random entity corruption, independently applied to both head and tail entities. During evaluation, the filtered ranking protocol was applied, excluding all known positive triples from the negative candidate pool before computing ranking metrics. Final ranking metrics were computed as the mean of head prediction and tail prediction results. All models were evaluated across three independent runs, with mean and standard deviation (SD) reported throughout. All models were trained and evaluated on an NVIDIA A100 80 GB PCIe GPU. Additional configurations are detailed in Supplementary Section F. Performance was evaluated using multiple metrics. For link prediction, we used Mean Rank (MR) (lower is better), Mean Reciprocal Rank (MRR) (higher is better), and Hits@k ($k = \{1, 3, 10\}$, higher is better), which measures the proportion of correct predictions in the top k results. MR measures the average rank of correct entities, MRR emphasises top-ranked predictions, and Hits@k assesses prediction accuracy at different cutoffs. For triple classification tasks, we used the Area Under the Receiver Operating Characteristic Curve (AUROC) and the Area Under the Precision–Recall Curve (AUPRC). AUROC measures the model's ability to distinguish between positive and negative examples across all classification thresholds, whilst AUPRC is particularly informative for imbalanced datasets by focusing on precision–recall trade-offs. Detailed mathematical formulations for all metrics are provided in Supplementary Section H.

2.6. Biomarker prediction and validation

The best-performing link prediction model was applied to score all entities in the KG based on predicted relationship strength to preterm birth, gestational diabetes, and preeclampsia. Relationships were classified as negative correlation (protective factors) or positive correlation (risk factors). Predictions were categorised as known (present in training data) or novel (absent from training data).

Known clinical interventions for the three complications were extracted from Cochrane Library systematic reviews [57], a curated database of systematic reviews considered the gold standard for evidence-based clinical knowledge, and compared against model predictions to assess recovery of established interventions. The utility of the KG for biological pathway exploration was evaluated by testing whether semantic chain traversal could reconstruct known pregnancy pathways [58]. The corticotropin-releasing hormone (CRH) pathway, which controls pregnancy timing and labour onset, was selected as a illustrative case study. Starting from CRH as the initial query, the system traced multi-hop relationships through the maternal, placental, and fetal compartments to identify components of the hypothalamic–pituitary–adrenal axis and associated feedback loops.

Novel predictions absent from training data were cross-referenced against genetic associations confirmed by searching the genome-wide association studies (GWAS) Catalogue [59] for reported variants associated with pregnancy complications and related traits. Tissue-specific expression was verified using the Human Protein Atlas [60] to confirm expression in pregnancy-relevant tissues. Biological mechanisms were assessed by examining receptor function, downstream signalling, and published experimental evidence, including animal models. Novel predictions were ranked by combined score. Generic clinical terms and pregnancy-related outcome entities were excluded, and the highest-ranked biologically specific candidates were cross-referenced against GWAS catalogues.

3. Results

The results follow the pipeline from construction to evaluation, covering KG statistics (Section 3.1), entity embeddings and structural patterns (Section 3.2), LLM benchmarking (Section 3.3), link prediction performance (Section 3.4), clinical evidence recovery, biomarker predictions (Section 3.5), and the web interface (Section 3.6).

3.1. PregKG statistics

PregKG is the graph that PregBase is built on. Initial extraction from 8420 pregnancy-related articles (5137 full-texts and 3283 abstracts) between 1960 and 2025 yielded 531,175 triples. After quality filtering (removing self-loops, invalid entities, and missing fields), all triples were validated through UMLS ontology mapping and assigned to three confidence tiers (Table 2). Entity mapping using SapPMBERT [61] achieved approximately 10% higher accuracy and produced more consistent alignments than Flan-T5 [62], consistent with our previous

Table 2
Triple counts at each stage of the PregKG construction pipeline, from initial extraction through tiered validation to the final knowledge graph.

Stage	Triples
Extracted	531,175
Tier 1	6039
Tier 2	177,466
Tier 3 (excluded)	347,670
Tier 1 after filtering	5373
Tier 2 after scoring (top 30%)	58,714
Final PregKG	**64,087**

work [48]. The hybrid GNN (GAT + R-GCN + GCN) trained on Tier 1 achieved a clear separation between high- and low-confidence triples compared to the baseline R-GCN alone (see Supplementary Figure S3). PMI scores for entity co-occurrence were computed across the literature corpus. Through CPU-based parallel processing, PMI computation was reduced from 300 to 15 h (see Supplementary Section E). We combined the hybrid model's graph scores with PMI scores using a $0.7 : 0.3$ weighting. A sensitivity analysis across α values from 0.5 to 0.9 confirmed that tier separation remained highly significant at all values (all $t > 109$, $p < 0.001$; unpaired two-sample t-test, $n = 6039$ Tier 1 and $n = 177{,}466$ Tier 2 triples; Supplementary Table S4). The results were consistent across the full range, with $\alpha = 0.7$ selected to reflect the greater reliability of the domain-trained graph signal. Applying top-30% percentile selection on combined scores (range 0.03–1.00, mean 0.69), 58,714 validated Tier 2 triples were retained at a score cutoff of 0.7617, which exceeds the Tier 2 mean of 0.6845. The 30% threshold is a deliberate design decision. A stricter cutoff retains fewer triples at higher confidence, while a more permissive cutoff increases coverage but admits lower-confidence relationships. The 30% was selected as the balance point, and the full tradeoff across thresholds from 20% to 40% is shown in Supplementary Table S5.

The final PregKG contains **64,087** high-confidence triples (5373 Tier 1 + 58,714 validated Tier 2) with **13,303** distinct biomedical entities connected by **8** relationship types, spanning **155** semantic types and **50** vocabulary sources. The most frequent relation type is Association (52,505 triples), followed by Positive_Correlation (7269) and Negative_Correlation (1953); the full distribution is shown in Supplementary Figure S9. The clustering coefficient (0.143) is 286-fold higher than that of a degree-matched random graph (0.001), and the mean shortest path length within the largest connected component is 2.94 hops, with a broad degree distribution following a power law ($R^2 = 0.976$). Nested triple construction produced 3,723,951 two-hop semantic chains across 7827 unique outer subjects and 9042 unique inner objects. Detailed statistics are presented in Supplementary Figures S4-S11.

3.2. PregKG entity embeddings reveal structural patterns

To analyse the distribution of the learned entities in PregBase, we conducted a visualisation, where the embeddings of the entities were projected into two dimensions using Uniform Manifold Approximation and Projection (UMAP) (*n_neighbors=15, min_dist=0.1, random_state=42*) [63], as shown in Fig. 3. It was found that related concepts group together spatially in the UMAP. For instance, pathologic functions, including preterm birth, cluster in the left region, while pregnancy-related conditions and outcomes, such as preeclampsia and miscarriage, form clusters closer to preterm birth. On the opposite spectrum of the UMAP lie the pharmacologic substances, forming a distinct cluster on the right, where molecular markers like leptin and IGF-I appear alongside interventions such as ritodrine and misoprostol, whereas anatomical features such as the myometrium group lie in the lower centre. Diseases are distributed throughout and are positioned near related pathologic functions. To quantify the observed clustering, we applied k-means (k=7) to the original high-dimensional embeddings across the seven focus semantic categories, yielding a cluster purity of 0.65. Annotated entities represent the highest-frequency examples from each semantic category. The observed semantic clustering confirms that the entity embeddings capture meaningful biomedical structure, which supports the downstream tasks in Sections 3.4 and 3.5.

3.3. Ensemble shuffle prompting achieves the most optimal LLM-based triple extraction

We compared different prompting strategies for LLM-based triplet extraction in our PregKG. Among the evaluated prompting strategies, ensemble shuffle consistently achieved the highest macro-average F1 scores across the majority of models (Fig. 4) on the BioRED [45] and BioREL [47] datasets. Ensemble shuffle randomly reorders example triples across ten independent runs and retains only predictions confirmed by majority vote, reducing position bias inherent in fixed-order prompting. It performed best in 11 out of 13 LLM models (Supplementary Figure S1). Average precision and recall alongside F1 for all 13 models are reported in Supplementary Table S3. While ensemble prompting achieved the best isolated performance, we did not restrict our pipeline to a single strategy. This ensured that complementary triples extracted by different strategies were preserved and unified.

The top-performing models overall were GPT-4o and MedGemma-27B, both reaching an F1 score of 0.815 under ensemble prompting. In biomedical triple extraction, precision is as critical as recall since low precision leads to spurious triples propagating through the pipeline. MedGemma-27B achieved an average precision of 0.870 and a recall of 0.783 under ensemble shuffle (Supplementary Table S3). A confusion matrix for relation extraction under ensemble shuffle is provided in Supplementary Figure S2. Other high-performing open-source models included Q LLaMA3.3 70B, LLaMA3.3 70B, and LLaMA3.1 8B, each scoring above 0.77. MedGemma-27B was selected as the final model for the pipeline due to its strong overall performance and open-source availability. MedGemma-27B was re-evaluated under ensemble shuffle across three independent runs, achieving macro-average F1 = 0.825 ± 0.001, Precision = 0.864 ± 0.007, and Recall = 0.805 ± 0.004.

3.4. Link prediction on PregKG discovers new connections between pregnancy-related concepts

We assessed whether PregKG captures meaningful biomedical relationships by testing its ability to predict missing links between pregnancy-related concepts. This capability reflects a key strength of any KG, which is uncovering novel associations beyond curated data. Evaluating link prediction thus provides a direct measure of PregBase's potential to generate new biological or clinical insights.

To establish benchmarks, we first applied standard KG completion models. Embedding-based methods such as TransE [52], DistMult [53], and ComplEx [54] learn geometric representations of entities and relations to infer plausible connections. We used MR, MRR and Hits@k (where k=1, 3, 10). For triplet classification and validation tasks, we used AUROC and AUPRC. TransE achieved an MRR of 0.281 ± 0.005 with Hits@10 of 0.613 ± 0.004, while DistMult and ComplEx reached MRRs of 0.481 ± 0.015 and 0.459 ± 0.003, respectively, indicating moderate predictive ability.

We next evaluated GNN models, which leverage structural context through message passing. CompGCN showed limited accuracy (MRR = 0.180 ± 0.072), whereas R-GCN performed substantially better (MRR = 0.564 ± 0.003; Hits@10 = 0.890 ± 0.004; AUROC = 0.84; AUPRC = 0.84), highlighting the value of relational aggregation reasoning but also the expressiveness constraints of standard GNN formulations (Table 3, Fig. 5).

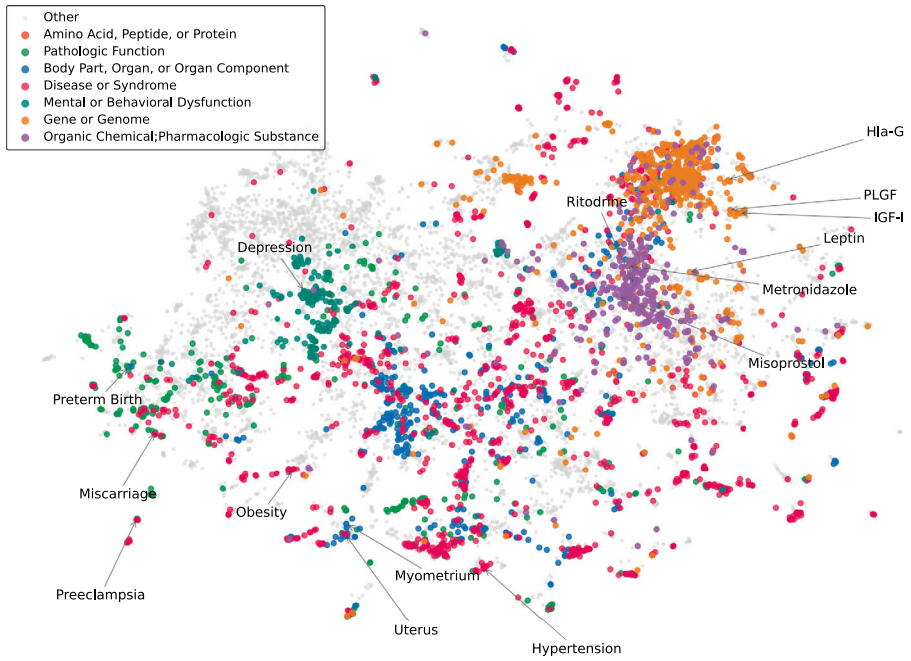


Fig. 3. UMAP visualisation of entity embeddings (n_neighbors=15, min_dist=0.1; cluster purity=0.65). Each point represents an entity, with colours indicating semantic type as shown in the legend. Entities with similar biological meanings cluster together closely, while the entities that are not related to each other are far away from each other. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

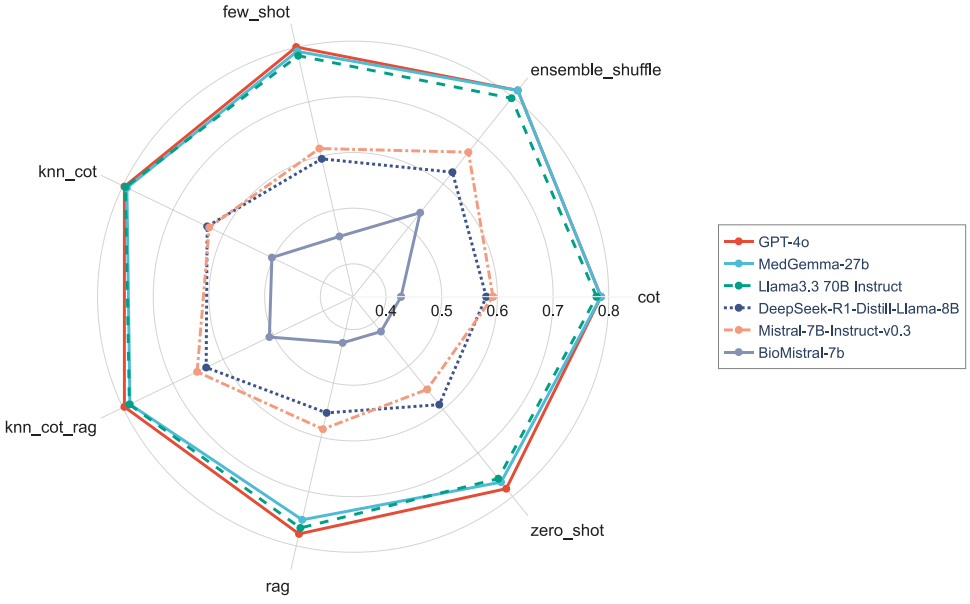


Fig. 4. Benchmarking performance in terms of F1-score for different prompting strategies and LLM models. Spider chart showing macro-average F1 scores across entity recognition, relation extraction, and type identification tasks. Each axis represents a prompting strategy, and each coloured line represents a different LLM model. Distance from the centre indicates F1 score performance, with values closer to the outer edge indicating better performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To overcome these limits, we developed hybrid architectures that integrate embedding-based scoring with GNN-derived relational features. This design enables joint exploitation of local connectivity and global semantics. The hybrid models consistently surpassed all baselines across three independent runs (Table 3, Fig. 5). V1-DistMult achieved the highest ranking accuracy (MRR = 0.853 ± 0.023; Hits@1 = 0.778 ± 0.030). V2-Edge-DistMult delivered near-perfect retrieval (Hits@10 = 0.982 ± 0.011; Hits@3 = 0.961 ± 0.013; MR = 1.470 ±0.085) and the highest AUROC = 0.96, maintaining precision across all recall levels. A direct comparison between V2-Edge-DistMult and R-GCN confirmed a 48.2% MRR improvement (MRR = 0.836 ± 0.022 vs 0.564 ± 0.003; $t = 20.84$, $p < 0.001$), demonstrating that the observed performance gains are statistically significant and not attributable to stochastic variation in training. The 95% confidence intervals confirm non-overlapping performance ranges between the best model and baseline (V2-Edge-DistMult MRR: 95% CI 0.781–0.892; R-GCN MRR: 95% CI 0.557–0.572). Other hybrids, including V1-ConvKB and V2-ConvKB (Combined Loss), also achieved strong discrimination (AUROC = 0.93–0.94).

Table 3
Link prediction performance of embedding-based models (TransE, DistMult, ComplEX), GNN-based baseline link prediction models (CompGCN, R-GCN), and hybrid models introduced in this work. Mean Rank (MR), Mean Reciprocal Rank (MRR), Hits@K, and inference time per epoch are reported as mean ± SD across three independent runs.

Model	MR↓	MRR↑	Hits@1↑	Hits@3↑	Hits@10↑	Time (s)
TransE	8.864 ± 0.028	0.281 ± 0.005	0.139 ± 0.002	0.291 ± 0.001	0.613 ± 0.004	2.48
DistMult	7.105 ± 0.034	0.481 ± 0.015	0.344 ± 0.019	0.550 ± 0.016	0.716 ± 0.005	0.44
ComplEX	7.138 ± 0.059	0.459 ± 0.003	0.319 ± 0.003	0.524 ± 0.002	0.705 ± 0.004	3.70
CompGCN	13.768 ± 1.112	0.180 ± 0.072	0.125 ± 0.046	0.197 ± 0.051	0.314 ± 0.078	13.72
R-GCN	4.141 ± 0.042	0.564 ± 0.003	0.411 ± 0.003	0.644 ± 0.004	0.890 ± 0.004	6.57
V1-ConvKB	9.513 ± 0.582	0.381 ± 0.032	0.233 ± 0.047	0.466 ± 0.034	0.590 ± 0.055	9.02
V1-DistMult	2.148 ± 0.260	**0.853 ± 0.023**	**0.778 ± 0.030**	0.920 ± 0.018	0.956 ± 0.011	7.46
V2-ConvKB (Combined Loss)	12.470 ± 1.234	0.291 ± 0.040	0.174 ± 0.028	0.337 ± 0.053	0.414 ± 0.086	8.15
V2-DistMult (Combined Loss)	2.532 ± 0.898	0.803 ± 0.016	0.700 ± 0.010	0.902 ± 0.049	0.936 ± 0.053	7.24
V2-Edge-ConvKB	7.031 ± 0.679	0.469 ± 0.013	0.301 ± 0.010	0.586 ± 0.030	0.732 ± 0.039	8.45
V2-Edge-DistMult	**1.470 ± 0.085**	0.836 ± 0.022	0.716 ± 0.034	**0.961 ± 0.013**	**0.982 ± 0.011**	7.30

*V1 architectures use GAT without relation awareness + RGCN + GCN residual. V2 architectures use GAT relation-aware + RGCN + GCN residual. Results reported as mean ± SD across three independent runs (seeds 42, 123, 456). Best performance is in bold.

Table 4
Recovery of Cochrane Evidence-Based Interventions. Extracted interventions were present as direct triples in PregKG; inferred interventions were identified through link prediction.

Condition	Cochrane total	Total recovered	Recovery rate (%)	Extracted	Inferred
Preterm Birth	375	348	92.8	146	202
Gestational Diabetes	104	97	93.3	29	68
Preeclampsia	184	123	66.8	4	119

Together, these results demonstrate that combining relational reasoning with embedding-based inference markedly improved PregBase's ability to uncover hidden biomedical relationships. The hybrid architectures capture the complex, multi-relational structure of pregnancy knowledge far more effectively than conventional approaches. Supporting ablation analyses are provided in the Supplementary Table S3 and Section F.

3.5. PregBase enables clinical evidence recovery and novel biomarker prediction

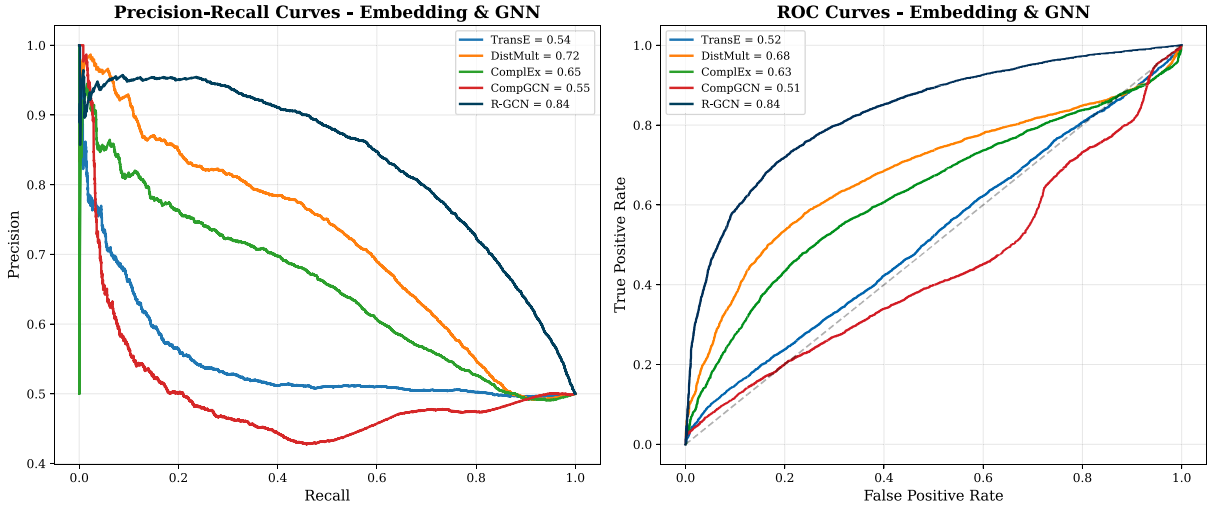
Clinical Evidence Recovery: Cochrane reviews are the gold standard summaries of clinical trial evidence [64,65]. We conducted a test to see whether the link prediction model captures real clinical knowledge from PregKG by comparing predictions against Cochrane systematic reviews. We extracted all documented interventions for preterm birth, gestational diabetes, and preeclampsia from Cochrane databases. Then, we checked how many of these interventions appeared in our model's predictions. An intervention was considered recovered if the corresponding entity appeared in model predictions, identified through exact substring matching, keyword matching, or fuzzy string matching (similarity threshold 0.80). Where multiple matches were found, the highest-scoring prediction was retained. The hybrid model achieved strong recovery rates *(Recovered interventions/Cochrane total) × 100)* across all three pregnancy complications, with an average of 84.3% (Table 4). For preterm birth, 146 interventions were directly extracted from the literature, and 202 were inferred through link prediction. For gestational diabetes, 29 were extracted and 68 inferred. Preeclampsia showed a lower overall recovery rate (66.8%), with the majority of recovered interventions inferred (119 of 123).

Validating Known Biomarkers: Predictions were grouped into known biomarkers (present in the training KG) and novel biomarkers (absent from the training data). The best-performing link prediction model was applied to identify biomarkers for the three pregnancy complications systematically. Every entity in the KG was scored based on predicted relationship strength to a biomarker. Relationships were further classified as negative correlation (protective factors) or positive correlation (risk factors). Representative known biomarkers are shown in Table 5. For preterm birth, protective interventions like progesterone received high scores with negative correlations. It is a clinical treatment deemed to prevent preterm birth [66]. Risk factors like maternal infections and smoking showed positive correlations, matching clinical observations [67–69]. Vitamin B12 deficiency also showed a positive correlation with preterm birth, consistent with its role in increasing homocysteine

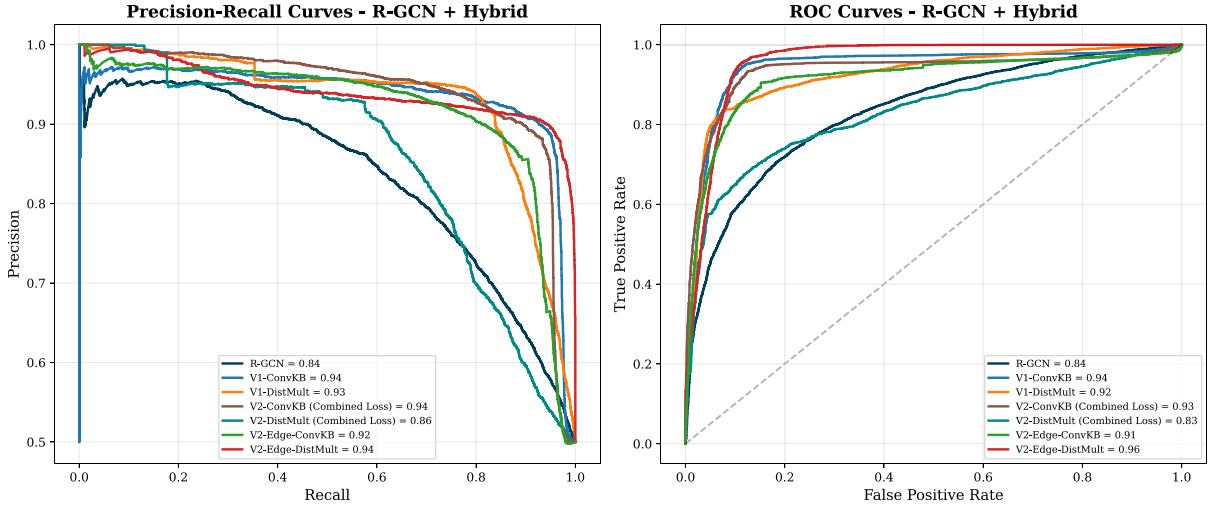
Table 5
Predicted Known Biomarkers for Three Pregnancy Complications: Preterm Birth, Gestational Diabetes and Preeclampsia.

Condition	Biomarker	Relationship type
Preterm Birth	Magnesium sulphate	Negative (Protective)
	Progesterone	Negative (Protective)
	Smoking	Positive (Risk)
	ISG15	Positive (Risk)
	IL-18	Positive (Risk)
	Apolipoprotein B	Positive (Risk)
	Vitamin B12 deficiency	Positive (Risk)
	Atopobium vaginae	Positive (Risk)
	Low Socioeconomic Status	Positive (Risk)
	Mono-benzyl phthalate	Positive (Risk)
Gestational Diabetes	Body mass index (BMI)	Positive (Risk)
	Insulin resistance	Positive (Risk)
	Polycystic ovary syndrome (PCOS)	Positive (Risk)
	Leptin	Positive (Risk)
	Adiponectin	Negative (Protective)
	Tumour necrosis factor-alpha (TNF-α)	Positive (Risk)
	Placental growth factor (PlGF)	Positive (Risk)
	Insulin-like growth factor-1 (IGF-1)	Positive (Risk)
	Mediterranean diet adherence	Negative (Protective)
	Metformin use	Negative (Protective)
Preeclampsia	Hypertension	Positive (Risk)
	Periodontal disease	Positive (Risk)
	Polycystic ovary syndrome (PCOS)	Positive (Risk)
	Oxidative stress	Positive (Risk)
	Air pollution	Positive (Risk)
	Overweight	Positive (Risk)
	SARS-CoV-2 infection	Positive (Risk)
	Magnesium sulphate	Negative (Protective)
	Smoking*	Negative (Protective)
	Placental growth factor (PlGF)	Negative (Protective)

*Although smoking appears as negatively correlated (protective), this reflects the well-documented epidemiological paradox where smoking is associated with reduced preeclampsia incidence, likely due to altered angiogenic pathways, and should *not* be interpreted as a clinical recommendation.



(a) ROC and PRC for embedding-based and baseline GNNs



(b) ROC and PRC for baseline RGCN and hybrid models

Fig. 5. ROC and Precision-Recall curves for baseline and hybrid models computed from a representative seed run. (a) Baseline R-GCN with the highest AUROC/AUPR: 0.84. (b) Hybrid models outperform, with V2-Edge-DistMult reaching AUROC 0.96.

and oxidative stress [69]. Similar patterns emerged for gestational diabetes and preeclampsia. As validated in clinical studies, dietary modification appeared as a protective factor for gestational diabetes [70], while body mass index (BMI) and polycystic ovary syndrome (PCOS) were identified as risks [71,72]. For preeclampsia, magnesium sulphate showed a negative correlation [73], whereas periodontal disease and air pollution showed positive correlations [74–77]. Smoking was negatively correlated with preeclampsia, though this reflects a known epidemiological paradox and should not be interpreted as protective [78].

Novel Biomarker Predictions: Beyond known relationships, the model generated novel computationally derived predictions for each condition: 25,959 for preterm birth, 10,197 for gestational diabetes, and 2222 for preeclampsia. After excluding generic clinical terms, the highest-ranked biologically specific candidates were cross-referenced against GWAS Catalogue [59] and the Human Protein Atlas [60]; the rank distribution of all candidates relative to reported ones is shown in Supplementary Figure S13. The top predictions across the three pregnancy conditions are presented in Table 6.

Independent genomic evidence was obtained through significant gene-level associations in the GWAS Catalogue [59], where p-values represent GWAS association statistics. For example, multiple variants within *YAP1* gene such as rs1894116-G ($p = 1 \times 10^{-22}$), rs11225154-A ($p = 8 \times 10^{-11}$), and rs10895278-C ($p = 7 \times 10^{-13}$) were significantly associated with endocrine and reproductive traits such as polycystic ovary syndrome, sex hormone-binding globulin levels, and offspring birth weight. These findings support Yes-associated protein (*YAP1*) as a biologically relevant gene involved in hormonal and placental regulation. This is consistent with the prediction as a novel candidate (not present in the training set) linking pregnancy-related disorders.

To see whether novel predictions were aligned with biological ground truth, we examined Prostaglandin F2α (PGF2α), one of the predicted biomarkers for preterm birth [79,80]. The receptor for PGF2α is encoded by the *PTGFR* gene. The Human Protein Atlas [60] confirmed strong expression of this receptor in pregnancy-related tissues, including endometrium, cervix, placenta, and ovary (see Supplementary Figure S11). The receptor mediates smooth muscle contraction and luteolysis. Experimental studies in mice have shown that activation of this receptor triggers parturition in ovarian luteal cells, supporting its role in labour onset [79,81]. This matches the predicted role in labour regulation and supports the biological plausibility of the model's prediction.

Table 6
Top Novel Biomarker Predictions Across Three Conditions: Preterm Birth, Gestational Diabetes and Preeclampsia.

Condition	Biomarker	Relationship type
Preterm birth	ISG15	Positive (Risk)
	N-acetylcysteine (NAC)	Negative (Protective)
	YAP1	Positive (Risk)
	CRH	Positive (Risk)
	Mono(2-ethylhexyl) phthalate	Positive (Risk)
	Prostaglandin F2alpha (PGF2α)	Positive (Risk)
Gestational Diabetes	CCL2	Positive (Risk)
	VEGF	Positive (Risk)
	IL-6	Positive (Risk)
	Thyroxine (FT4)	Negative (Protective)
	TCF7L2	Positive (Risk)
	N6-methyladenosine (m^6A RNA methylation)	Negative (Protective)
Preeclampsia	Gut Microbiota	Positive (Risk)
	Leptin	Positive (Risk)
	VEGF	Positive (Risk)
	CRH	Positive (Risk)
	Racism	Positive (Risk)
	Dexamethasone	Negative (Protective)

The utility of the PregBase was presented through a case study of semantic chain exploration in the CRH pathway, which controls pregnancy timing and the start of labour [58]. The automated reasoning system correctly identified several parts of the maternal–fetal hormonal feedback loop through semantic chain traversal (Fig. 6). Starting with CRH as the root entity, the system explored 411 unique semantic chains with lengths ranging from 1 to 5 triples, predicting links involving the uterus and prostaglandins (Fig. 6A). The 20 most traversed connections (Fig. 6B) show CRH driving preterm labour through prostaglandin production and uterine contractions. This focused view reveals the hub-and-spoke architecture of the pathway, with CRH acting as the central initiator and prostaglandins serving as the key intermediate linking hormonal signals to mechanical uterine activity.

3.6. PregBase's web interface enables interactive search and visualisation

PregBase is accessible through an interactive web interface.[1] The platform displays the KG as a network visualisation with colour-coded entities and relationships. Nodes carry UMLS semantic type and vocabulary source as attributes, and edges carry relation type, confidence tier, and modifier. The query panel supports direct relationship searches between entities (e.g., *"progesterone"* and *"preterm birth"*), filtering by relationship type, semantic type, or vocabulary, multi-hop semantic chain tracing, and entity-focused neighbourhood exploration with focus mode. Users can download query results in CSV format. An integrated virtual AI chatbot accepts conversational queries in natural language like entity-focused searches (*"Focus on progesterone"*), relation filtering (*"Which entities have Association relationship?"*), semantic chain exploration, relationship discovery (*"Show connection between preterm birth and inflammation"*), topic summaries, and visual analytics (word clouds). The chatbot parses natural language queries using rule-based intent detection to identify query type and extract entities, returning results directly. The interface is built on a NetworkX [82] directed property graph and deployed through *vis.js* [83]. A screenshot of the interactive interface is provided in the Supplementary Figure S12.

4. Discussion

4.1. Key implications

PregBase is the first interactive knowledge base specifically designed for pregnancy research. Prompt design had a substantial effect on extraction quality. *Ensemble shuffle prompting* achieved strong performance without fine-tuning by combining multiple reasoning patterns within a single prompt. Whether this transfers to other biomedical domains remains to be tested, though a prompt-only approach requires no domain-specific training data.

Flat triples lose the sequential order in which biological events unfold. The nested triple structure captures this directly. For example, while the flat triples *(Cortisol, Negative_Correlation, CRH)* and *(CRH, Positive_Correlation, Prostaglandin Production)* exist independently in PregKG, the nested structure *(Cortisol, Negative_Correlation, (CRH, Positive_Correlation, Prostaglandin Production))* encodes the full feedback chain in a single queryable statement, showing that cortisol suppresses CRH which in turn drives prostaglandin-mediated labour onset. To balance reliability and coverage, extracted triples were organised into a three-tier classification system based on mapping confidence and statistical scoring. High-confidence triples support direct use, whilst lower-confidence triples preserve potentially informative signals that may generate new hypotheses. Despite comprising a small fraction of extracted triples, Tier 1 covers all 8 relation types and 142 of the 151 UMLS semantic types present in Tier 2, providing the structural diversity needed to train a generalisable scoring model. Nested structures and tiered classification together allow researchers to trace complex biological mechanisms at appropriate confidence levels. The graph topology supports this, with hub entities bridging disparate biological processes and short average path lengths supporting multi-hop traversal across the knowledge structure. The CRH pathway exemplifies this, where hub-and-spoke organisation emerges from the literature without explicit encoding. Such emergent structure suggests that the framework can recover system-level organisation from unstructured literature, a capability that extends broadly to other complex biomedical domains.

4.2. Clinical relevance

PregBase predictions aligned with established clinical evidence. Link prediction recovered known interventions from Cochrane systematic reviews [57]. Protective factors and risk markers were among the recovered associations. Unlike edge classification, which requires labelled positive and negative edges, link prediction is better suited to an incomplete literature-derived KG where absent edges do not reliably indicate true negatives [84]. The filtered ranking protocol handles this open-world assumption directly, and the formulation aligns naturally with surfacing novel biological associations rather than classifying known edge types. Progesterone showed a negative correlation with preterm birth, consistent with clinical evidence that vaginal progesterone reduces preterm birth rates by 45% in women with a short cervix [85]. Smoking was correctly identified as a risk factor, confirmed by

[1] https://pregknowledgebase.com

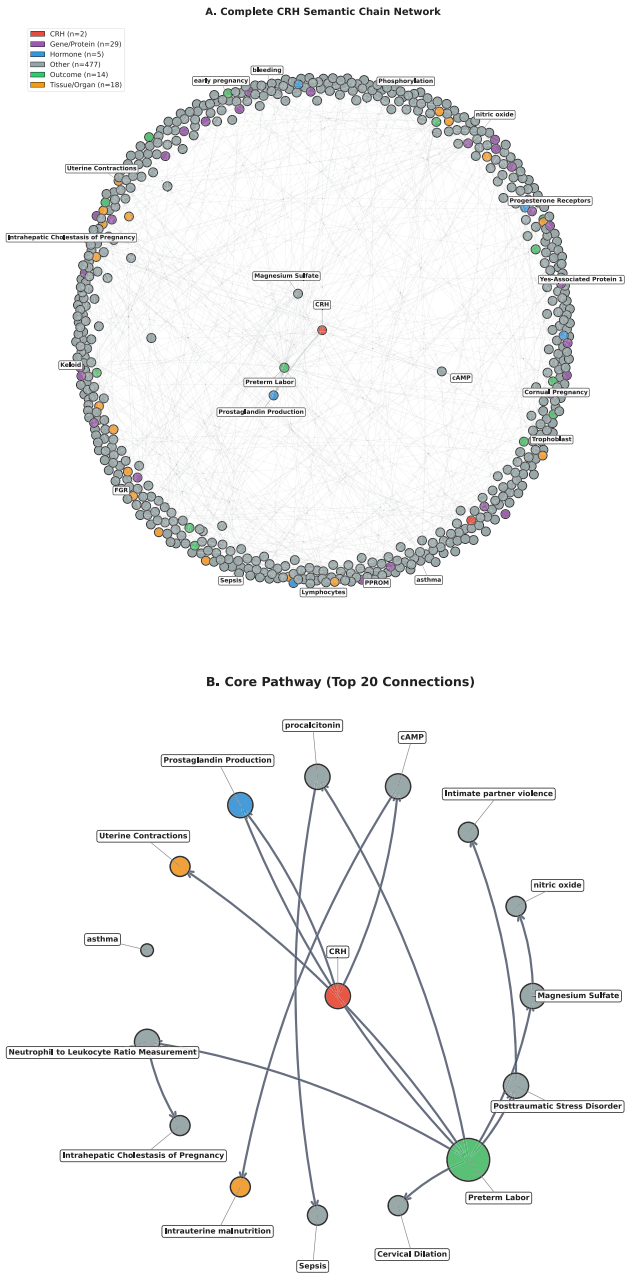


Fig. 6. CRH network analysis case study. **(A)** shows the complete CRH semantic chain network with 411 semantic chains originating from CRH. The circular layout emphasises CRH's central role in initiating downstream biological cascades leading to labour onset. Most chains (>75%) contain 4–5 triples, demonstrating extensive multi-hop reasoning beyond direct relationships. **(B)** shows the core CRH pathway network with the subgraph showing the top 20 strongest connections by edge weight in the chain. CRH (red, centre) connects to preterm labour (green) through prostaglandin production (blue), representing the primary cascade. Additional critical intermediates include magnesium sulphate (a tocolytic agent), cAMP, and uterine contractions (orange). The hierarchical radial layout positions CRH at the centre, with directly connected entities arranged in a circular pattern. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

meta-analyses demonstrating 27% increased preterm birth risk among smokers, with increased risk evident even at low-intensity exposure of 1–2 cigarettes daily [86,87]. Insulin resistance showed a strong association with gestational diabetes, with studies reporting a 5.3-fold increased risk of hypertensive disorders and a 1.65-fold risk of large-for-gestational-age infants in women with insulin resistance [88]. This metabolic dysfunction is linked to polycystic ovary syndrome (PCOS), a disease entity showing 20.6% pooled gestational diabetes incidence with some cohorts approaching 40% [72,89]. Among hormonal biomarkers, adiponectin, an adipocyte-derived protein, showed predicted inverse correlations with concentrations below 10.29 mg/mL combined with elevated BMI increases gestational diabetes risk [90], whilst leptin, another adipokine, showed predicted positive associations confirmed by nearly doubled maternal serum concentrations in preeclampsia (45.6 vs 27.0 ng/mL) [91,92]. Magnesium sulphate was negatively correlated with preterm birth, which is consistent with clinical evidence that antenatal magnesium sulphate reduces adverse neonatal outcomes through fetal neuroprotection [93]. This correlation does not imply efficacy in preventing preterm labour [94], and predictions should be interpreted alongside clinical expertise and systematic review evidence.

Several predicted associations showed biological plausibility when cross-referenced against genetic and tissue-expression databases [59,60]. YAP1 [95] and prostaglandin F2α [79,80] are two examples where independent data sources support the predicted links. The system also identified racism as a predicted risk factor for preeclampsia, consistent with documented disparities where Black women with preeclampsia demonstrate 60% higher risk for severe morbidity compared to White women [96]. Chronic stress from racial discrimination has been linked to HPA axis dysregulation and inflammatory pathway activation [97,98], and the prediction reflects this. The KG captures both biological and social determinants of pregnancy complications. Socio-environmental and molecular factors are particularly relevant for health disparities research. In practice, a clinician querying CRH in PregBase receives 411 semantic chains across the maternal–fetal hormonal cascade, with prostaglandins identified as the key intermediate and magnesium sulphate as negatively correlated with the CRH pathway, consistent with its clinical use as a tocolytic agent (Fig. 6). Cervical length not appearing as a mechanistic intermediate reflects gaps in literature coverage rather than biological absence. Lower-confidence connections require verification against primary sources. The output does not replace clinical judgment, but points a clinician towards the relevant mechanistic pathways without manually searching thousands of papers. The accompanying web interface extends these capabilities directly to clinicians and researchers without programming expertise. By enabling intuitive exploration of multi-hop biological relationships, users can trace mechanistic chains, identify candidate biomarkers, and generate hypotheses that integrate evidence across thousands of studies. The extraction pipeline is also structured to support incremental literature integration as new findings emerge, working towards a long-standing need for current and computable knowledge in maternal health research.

4.3. Limitations and future work

Our developed combination of the KG and link-prediction models enables the discovery of previously unrecognised associations by analysing patterns across literature-derived relationships. These inferred links can generate valuable hypotheses but require careful interpretation. For example, the observed inverse association between smoking and preeclampsia reflects a well-documented epidemiological paradox where light to heavy smoking shows reduced odds ratios (0.66–0.51) for preeclampsia despite being harmful for other pregnancy outcomes [78], underscoring that our system reveals correlations present in the literature but does not serve as proof for causality. The lower Cochrane recovery rate for preeclampsia partly reflects imperfect entity normalisation. Although UMLS maps "pre-eclampsia" to the preferred term "preeclampsia", 14 triples retained the hyphenated form, reducing matching coverage for Cochrane interventions using

this spelling. Additionally, preeclampsia-specific interventions are underrepresented in the corpus relative to preterm birth and gestational diabetes, with only 4 of 123 recovered interventions present as directly extracted triples. As with all approaches grounded in observational data, distinguishing genuine biological mechanisms from confounding remains a fundamental limitation. Extracted knowledge was also evaluated for three pregnancy complications only.

Errors in text extraction or entity mapping may also influence results, particularly in high-stakes clinical contexts. Manual inspection of 100 randomly sampled Tier 3 triples showed that 89% failed due to low entity similarity or vague process terms absent from UMLS, 10% used relation types outside the defined schema, and 1% had truncated entity strings. Most discarded triples were rejected by the UMLS validator rather than being incorrectly extracted. Knowledge was generated using a single model and, despite comprehensive benchmarking, systematic errors such as hallucinated entities cannot be fully ruled out. Further analysis of low-scoring predictions identified spurious associations, including *Sacroiliac Joint* linked to *preeclampsia* and *parenthood* correlated with *preterm birth*, driven by co-occurrence artefacts in the literature corpus rather than biological signal. A formal false discovery rate cannot be estimated without experimental ground truth for novel predictions. Predicted links reflect statistical associations in the literature and do not imply causality. Expert review and iterative refinement, therefore, remain integral to responsible use. The pipeline has not undergone formal human expert evaluation on pregnancy-specific text, which remains a priority for future work, as does formal evaluation of the chatbot component. The corpus also spans over six decades of shifting clinical terminology, which UMLS normalisation only partially resolves. Full-text extraction was limited to PMC open-access articles, which may not fully represent the breadth of pregnancy literature. The corpus is drawn from PubMed, which overrepresents English-language research and may underrepresent findings from low- and middle-income countries where pregnancy complications are most prevalent. Extracted triples are treated equally regardless of publication date. Older articles may contain outdated or superseded findings that have since been contradicted. Nonetheless, the modular architecture of the pipeline allows easy adaptation to new datasets, biomedical domains, and analytic tasks. Formal quantitative evaluation of nested triples in downstream tasks remains future work. The structured representation of pregnancy knowledge created here establishes a scalable foundation for broader applications, including automated synthesis in other areas of maternal and reproductive health.

The web interface is maintained on a six-monthly update cycle as the platform matures. Future development will focus on integrating external evidence sources, such as electronic health records, clinical trial registries, and patient-level data, to enable more personalised and clinically relevant analyses. Continuous ingestion of emerging literature and validation against independent datasets will further enhance robustness. Additionally, formal user studies with clinicians and researchers are needed to validate how the interactive interface supports practitioners in accessing and utilising the knowledge base for clinical decision-making and research workflows. These efforts will maintain PregBase's role as a transparent, research-oriented resource that bridges discovery and clinical translation.

5. Conclusion

We introduced PregBase, a comprehensive knowledge system that provides automated literature extraction, knowledge validation, link prediction, pathway reconstruction, biomarker prediction, and interactive exploration through natural language querying. To our knowledge, PregBase is the first hybrid and interactive knowledge system specifically designed for pregnancy research. We believe PregBase will offer a clear base for understanding how biological signals, clinical findings and risk factors interact across maternal outcomes. Our built KG gathers evidence from thousands of studies into one resource that reflects established biology and also reveals new patterns in the literature. In addition, our knowledge database creates the conditions needed for more systematic discovery in pregnancy research. Its modular design allows easy expansion with new data and new research needs. We believe our framework serves as an important step towards an integrated and data-driven understanding of pregnancy outcomes.

CRediT authorship contribution statement

Aashish Bhandari: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Deval Mehta:** Writing – review & editing, Methodology. **Karin Verspoor:** Writing – review & editing, Methodology. **Damiano Spina:** Writing – review & editing, Methodology. **Feng Xia:** Writing – review & editing, Supervision, Methodology. **Sonika Tyagi:** Writing – review & editing, Supervision, Resources, Methodology, Investigation, Conceptualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

AB acknowledges the RMIT STEM Scholarship, which supports his PhD research. The authors thank the RMIT AWS Cloud Supercomputing (RACE) Hub for computational resources, with particular thanks to Patrick Taylor for infrastructure setup and Mohammad Kazemi Beydokhti for early discussions that shaped the project's initial direction. The initial development phase was supported through RMIT's Enabling Impact Platform - RACE Demonstrator Competition Showcase 2024 (won the People's Choice Award).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.artmed.2026.103510.

Data availability

Code, model weights, and data can be accessed at https://gitlab.com/tyagilab/pregbase. The PregBase platform is available at https://pregknowledgebase.com.

References

[1] Walani Salimah R. Global burden of preterm birth. Int J Gynecol Obs 2020;150(1):31–3.
[2] Ye Wenrui, Luo Cong, Huang Jing, Li Chenglong, Liu Zhixiong, Liu Fangkun. Gestational diabetes mellitus and adverse pregnancy outcomes: systematic review and meta-analysis. BMJ 2022;377.
[3] Ghulmiyyah Labib, Sibai Baha. Maternal mortality from preeclampsia/eclampsia. Semin Perinatol 2012;36(1):56–9.
[4] Callaghan William M, Creanga Andreea A, Kuklina Elena V. Severe maternal morbidity among delivery and postpartum hospitalizations in the United States. Obs Gynecol 2012;120(5):1029–36.
[5] World Health Organization, et al. Trends in maternal mortality 2000 to 2017: estimates by WHO, UNICEF, UNFPA, world bank group and the united nations population division: executive summary. 2019.
[6] World Health Organization. Maternal deaths decline slowly with vast inequalities worldwide. 2019, https://www.who.int/news/item/19-09-2019-maternal-deaths-decline-slowly-with-vast-inequalities-worldwide. (Accessed: 11 October 2025).

[7] World Health Organization. Maternal mortality. 2025, https://www.who.int/news-room/fact-sheets/detail/maternal-mortality. (Accessed: 11 October 2025).

[8] Roberts James M, Heider Dominik, Bergman Lina, Thornburg Kent L. Vision for improving pregnancy health: innovation and the future of pregnancy research. Reprod Sci 2022;29(10):2908–20.

[9] Nyfløt Lill, Sitras Vasilis. Strategies to reduce global maternal mortality. 97, (6):2018, p. 639–40,

[10] Diana Sulis, Wahyuni Chatarina Umbul, Prasetyo Budi. Maternal complications and risk factors for mortality. J Public Health Res 2020;9(2).

[11] Zhang Yuan, Sui Xin, Pan Feng, Yu Kaixian, Li Keqiao, Tian Shubo, Erdengasileng Arslan, Han Qing, Wang Wanjing, Wang Jianan, et al. A comprehensive large-scale biomedical knowledge graph for AI-powered data-driven biomedical research. Nat Mach Intell 2025;1–13.

[12] Belsti Yitayeh, Moran Lisa, Mousa Aya, Goldstein Rebecca, Rolnik Daniel Lorber, Khomami Mahnaz Bahri, Kebede Mihiretu M, Teede Helena, Enticott Joanne. Analyzing electronic medical records to extract prepregnancy morbidities and pregnancy complications: Toward a learning health system. Learn Health Syst 2025;9(2):e10473.

[13] Foo Damien, Dunne Jennifer, Pereira Gavin, Gebremedhin Amanuel, Duko Bereket, Tessema Gizachew A. Diabetic and hypertensive disorders following miscarriage: a protocol for systematic review and meta-analysis. Int J Environ Res Public Health 2022;19(14):8324.

[14] Okoth Kelvin, Subramanian Anuradhaa, Chandan Joht Singh, Adderley Nicola J, Thomas G Neil, Nirantharakumar Krishnarajah, Antza Christina. Long term miscarriage-related hypertension and diabetes mellitus. evidence from a United Kingdom population-based cohort study. PLoS One 2022;17(1):e0261769.

[15] Rindflesch Thomas C, Fiszman Marcelo. SEMREP: A text mining tool for the biomedical literature. J Biomed Inform. 2003;36(6):439–50.

[16] Aronson Alan R, Lang François-Michel. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform. Assoc 2010;17(3):229–36.

[17] Lee Jinhyuk, Yoon Wonjin, Kim Sungdong, Kim Donghyeon, Kim Sunkyu, So Chan Ho, Kang Jaewoo. Biobert: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020;36(4):1234–40.

[18] Santosh Tokala Yaswanth Sri Sai, Chakraborty Prantika, Dutta Sudakshina, Sanyal Debarshi Kumar, Das Partha Pratim. Joint entity and relation extraction from scientific documents: Role of linguistic information and entity types. EEKE@ JCDL 2021;21:15–9.

[19] Gao Ting, Zhai Xue, Yang Chuan, Lv Linlin, Wang Han. Joint extraction of entity and relation based on fine-tuning BERT for long biomedical literatures. Bioinform Adv 2024;4(1):vbae194.

[20] Li Zhao, Wei Qiang, Huang Liang-Chin, Li Jianfu, Hu Yan, Chuang Yao-Shun, He Jianping, Das Avisha, Keloth Vipina Kuttichi, Yang Yuntao, et al. Ensemble pretrained language models to extract biomedical knowledge from literature. J Am Med Inform. Assoc 2024;31(9):1904–11.

[21] Yang Peiru, Wang Hongjun, Huang Yingzhuo, Yang Shuai, Zhang Ya, Huang Liang, Zhang Yuesong, Wang Guoxin, Yang Shizhong, He Liang, et al. LMKG: A large-scale and multi-source medical knowledge graph for intelligent medicine applications. Knowl-Based Syst 2024;284:111323.

[22] Kumari Madhavi, Chauhan Rohit, Garg Prabha. Medkg: enabling drug discovery through a unified biomedical knowledge graph. Mol Divers 2025;1–19.

[23] Gao Shan, Yu Kaixian, Yang Yue, Yu Sheng, Shi Chenglong, Wang Xueqin, Tang Niansheng, Zhu Hongtu. Large language model powered knowledge graph construction for mental health exploration. Nat Commun 2025;16(1):7526.

[24] Agrawal Monica, Hegselmann Stefan, Lang Hunter, Kim Yoon, Sontag David. Large language models are few-shot clinical information extractors. 2022, arXiv preprint arXiv:2205.12689.

[25] Nagar Aishik, Schlegel Viktor, Nguyen Thanh-Tung, Li Hao, Wu Yuping, Binici Kuluhan, Winkler Stefan. LLMs are not zero-shot reasoners for biomedical information extraction. Work Insights from Negat Results NLP 2025;106–20.

[26] Qi Xing, Yang Bo, Wang Shilong, Zhang Zhengping, Zhang Yucheng, Du Kaze. Few-shot and chain-of-thought prompting for equipment maintenance knowledge graph construction via large language models. Knowl-Based Syst 2026;115266.

[27] Sivarajkumar Sonish, Kelley Mark, Samolyk-Mazzanti Alyssa, Visweswaran Shyam, Wang Yanshan. An empirical evaluation of prompting strategies for large language models in zero-shot clinical natural language processing: algorithm development and validation study. JMIR Med Inform. 2024;12:e55318.

[28] Bai Xuefeng, He Song, Li Yi, Xie Yabo, Zhang Xin, Du Wenli, Li Jian-Rong. Construction of a knowledge graph for framework material enabled by large language models and its application. Npj Comput Mater 2025;11(1):51.

[29] Morris John H, Soman Karthik, Akbas Rabia E, Zhou Xiaoyuan, Smith Brett, Meng Elaine C, Huang Conrad C, Cerono Gabriel, Schenk Gundolf, Rizk-Jackson Angela, et al. The scalable precision medicine open knowledge engine (SPOKE): a massive knowledge graph of biomedical information. Bioinformatics 2023;39(2):btad080.

[30] Zhou Wei-Zhen, Li Wenke, Shen Huayan, Wang Ruby W, Chen Wen, Zhang Yujing, Zeng Qingyi, Wang Hao, Yuan Meng, Zeng Ziyi, et al. Chdbase: a comprehensive knowledgebase for congenital heart disease-related genes and clinical manifestations. Genom Proteom Bioinform 2023;21(1):216–27.

[31] ElShal Sarah, Tranchevent Léon-Charles, Sifrim Alejandro, Ardeshirdavani Amin, Davis Jesse, Moreau Yves. Beegle: from literature mining to disease-gene discovery. Nucleic Acids Res 2016;44(2):e18.

[32] Urdang Jacqueline G, Masters Stephanie, Edokobi Nneoma, Mukherjee Chitra, Quazi Arnib, Liem David A, Ahrens Monica, Wang Xuan, Whitham Megan. Text phrase-mining in identifying and classifying maternal proteins and genes across preeclampsia and similar pathologies. Physiol Rep 2025;13(6):e70262.

[33] Rodriguez Laritza M, Morrison Stephanie M, Greenberg Kathleen, Fushman Dina Demner. Mining the literature for genes associated with placenta-mediated maternal diseases. AMIA Annu Symp Proc 2018;2017:1498–507.

[34] Islam Muhammad Nazrul, Mustafina Sumaiya Nuha, Mahmud Tahasin, Khan Nafiz Imtiaz. Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda. BMC Pregnancy Childbirth 2022;22(1):348.

[35] Du Yuhan, McNestry Catherine, Wei Lan, Antoniadi Anna Markella, McAuliffe Fionnuala M, Mooney Catherine. Machine learning-based clinical decision support systems for pregnancy care: a systematic review. Int J Med Inform. 2023;173:105040.

[36] Chen Hsiang-Yang, Chuang Chao-Hua, Yang Yao-Jung, Wu Tung-Pi. Exploring the risk factors of preterm birth using data mining. Expert Syst Appl 2011;38(5):5384–7.

[37] Xiong Bo, Nayyeri Mojtaba, Luo Linhao, Wang Zihao, Pan Shirui, Staab Steffen. Neste: Modeling nested relational structures for knowledge graph reasoning. Proc the AAAI Conf Artif Intell 2024;38(8):9205–13.

[38] Pu Yiyuan, Beck Daniel, Verspoor Karin. Enriched knowledge representation in biological fields: a case study of literature-based discovery in alzheimer's disease. J Biomed Semant 2025;16(1):3.

[39] Chen Chih-Ping. Congenital malformations associated with maternal diabetes. Taiwan J Obs Gynecol 2005;44(1):1–7.

[40] Chandak Payal, Huang Kexin, Zitnik Marinka. Building a knowledge graph to enable precision medicine. Sci Data 2023;10(1):67.

[41] Shang Tianqi, Yang Shu, He Weiqing, Zhai Tianhua, Li Dawei, Hou Bojian, Chen Tianlong, Moore Jason H, Ritchie Marylyn D, Shen Li. Leveraging social determinants of health in alzheimer's research using llm-augmented literature mining and knowledge graphs. AMIA Summits Transl Sci Proc 2025;2025:491.

[42] Bodenreider Olivier. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32(suppl_1):D267–70.

[43] Xia Feng, Peng Ciyuan, Ren Jing, Febrinanto Falih Gozi, Luo Renqiang, Saikrishna Vidya, Yu Shuo, Kong Xiangjie. Graph learning. Found Trends Signal Process 2025;19(4):371–551. http://dx.doi.org/10.1561/2000000137.

[44] Cock Peter JA, Antao Tiago, Chang Jeffrey T, Chapman Brad A, Cox Cymon J, Dalke Andrew, Friedberg Iddo, Hamelryck Thomas, Kauff Frank, Wilczynski Bartek, et al. Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics 2009;25(11):1422.

[45] Luo Ling, Lai Po-Ting, Wei Chih-Hsuan, Arighi Cecilia N, Lu Zhiyong. Biored: a rich biomedical relation extraction dataset. Brief Bioinform 2022;23(5):bbac282.

[46] Unni Deepak R, Moxon Sierra AT, Bada Michael, Brush Matthew, Bruskiewich Richard, Caufield J Harry, Clemons Paul A, Dancik Vlado, Dumontier Michel, Fecho Karamarie, et al. Biolink model: A universal schema for knowledge graphs in clinical, biomedical, and translational science. Clin Transl Sci 2022;15(8):1848–55.

[47] Xing Rui, Luo Jie, Song Tengwei. BioRel: towards large-scale biomedical relation extraction. BMC Bioinformatics 2020;21(Suppl 16):543.

[48] Mainwood Samuel Thomas, Bhandari Aashish, Tyagi Sonika. Semantic encoding in medical llms for vocabulary standardisation. In: Proceedings of the 2025 18th Health Informatics Knowledge Management Conference. 2025, p. 1–7.

[49] Ankit Pal Malaikannan Sankarasubbu. OpenBioLLMs: Advancing open-source large language models for healthcare and life sciences. 2024, https://huggingface.co/aaditya/OpenBioLLM-Llama3-70B,

[50] Watford Sean M, Grashow Rachel G, Vanessa Y, Rudel Ruthann A, Friedman Katie Paul, Martin Matthew T. Novel application of normalized pointwise mutual information (NPMI) to mine biomedical literature for gene sets associated with disease: Use case in breast carcinogenesis. Comput Toxicol 2018;7:46–57.

[51] Xiao Yongkang, Hou Yu, Zhou Huixue, Diallo Gayo, Fiszman Marcelo, Wolfson Julian, Zhou Li, Kilicoglu Halil, Chen You, Su Chang, et al. Repurposing non-pharmacological interventions for alzheimer's disease through link prediction on biomedical literature. Sci Rep 2024;14(1):8693.

[52] Bordes Antoine, Usunier Nicolas, Garcia-Duran Alberto, Weston Jason, Yakhnenko Oksana. Translating embeddings for modeling multi-relational data. Adv Neural Inf Process Syst 2013;26.

[53] Yang Bishan, Yih Wen-tau, He Xiaodong, Gao Jianfeng, Deng Li. Embedding entities and relations for learning and inference in knowledge bases. 2014, arXiv preprint arXiv:1412.6575.

[54] Trouillon Théo, Welbl Johannes, Riedel Sebastian, Gaussier Éric, Bouchard Guillaume. Complex embeddings for simple link prediction. Int Conf Mach Learn 2016;2071–80.

[55] Vashishth Shikhar, Sanyal Soumya, Nitin Vikram, Talukdar Partha. Composition-based multi-relational graph convolutional networks. 2019, arXiv preprint arXiv:1911.03082.

[56] Schlichtkrull Michael, Kipf Thomas N, Bloem Peter, Van Den Berg Rianne, Titov Ivan, Welling Max. Modeling relational data with graph convolutional networks. Eur Semant Web Conf 2018;593–607.

[57] The Cochrane Collaboration. Cochrane library. 2025, https://www.cochranelibrary.com. (Accessed 12 September 2025).

[58] Herrera Christina L, Maiti Kaushik, Smith Roger. Preterm birth and corticotrophin-releasing hormone as a placental clock. Endocrinology 2023;164(2):bqac206.

[59] MacArthur Jacqueline, Bowler Emily, Cerezo Maria, Gil Laurent, Hall Peggy, Hastings Emma, Junkins Heather, McMahon Aoife, Milano Annalisa, Morales Joannella, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). Nucleic Acids Res 2017;45(D1):D896–901.

[60] Thul Peter J, Lindskog Cecilia. A subcellular map of the human proteome. Science 2017;356(6340):eaal3321. http://dx.doi.org/10.1126/science.aal3321.

[61] Liu Fangyu, Shareghi Ehsan, Meng Zaiqiao, Basaldella Marco, Collier Nigel. Self-alignment pretraining for biomedical entity representations. Proc the 2021 Conf the North Am Chapter of Assoc Comput Linguist: Hum Language Technol 2021;4228–38.

[62] Chung Hyung Won, Hou Le, Longpre Shayne, Zoph Barret, Tay Yi, Fedus William, Li Yunxuan, Wang Xuezhi, Dehghani Mostafa, Brahma Siddhartha, et al. Scaling instruction-finetuned language models. J Mach Learn Res 2024;25(70):1–53.

[63] McInnes Leland, Healy John, Melville James. Umap: Uniform manifold approximation and projection for dimension reduction. 2018, arXiv preprint arXiv:1802.03426.

[64] El Dib Regina P, Atallah Álvaro N, Andriolo Regis B. Mapping the cochrane evidence for decision making in health care. J Eval Clin Pract 2007;13(4):689–92.

[65] Starr Mark, Chalmers Iain, Clarke Mike, Oxman Andrew D. The origins, evolution, and future of the cochrane database of systematic reviews. Int J Technol Assess Health Care 2009;25(S1):182–95.

[66] Dodd Jodie M, Flenady Vicki J, Cincotta Robert, Crowther Caroline A. Progesterone for the prevention of preterm birth: a systematic review. Obs Gynecol 2008;112(1):127–34.

[67] Pararas MV, Skevaki CL, Kafetzis DA. Preterm birth due to maternal infection: causative pathogens and modes of prevention. Eur J Clin Microbiol Infect Dis 2006;25(9):562–9.

[68] Kyrklund-Blomberg Nina B, Granath Fredrik, Cnattingius Sven. Maternal smoking and causes of very preterm birth. Acta Obs et Gynecol Scand 2005;84(6):572–7.

[69] Park Bongsoo, Khanam Rasheda, Vinayachandran Vinesh, Baqui Abdullah H, London Stephanie J, Biswal Shyam. Epigenetic biomarkers and preterm birth. Environ Epigenetics 2020;6(1):dvaa005.

[70] Altemani Abdullah H, Alzaheb Riyadh A. The prevention of gestational diabetes mellitus (the role of lifestyle): a meta-analysis. Diabetol Metab Syndr 2022;14(1):83.

[71] Zhang Shuang, Liu Huikun, Li Nan, Dong Wei, Li Weiqin, Wang Leishen, Zhang Yu, Yang Yingzi, Leng Junhong. Relationship between gestational body mass index change and the risk of gestational diabetes mellitus: a community-based retrospective study of 41,845 pregnant women. BMC Pregnancy Childbirth 2022;22(1):336.

[72] Yan Qingzi, Qiu Dan, Liu Xiang, Xing Qichang, Liu Renzhu, Hu Yixiang. The incidence of gestational diabetes mellitus among women with polycystic ovary syndrome: a meta-analysis of longitudinal studies. BMC Pregnancy Childbirth 2022;22(1):370.

[73] Tukur Jamilu, Ahonsi Babatunde, Mohammed Ishaku Salisu, Araoyinbo Idowu, Okereke Ekechi, Babatunde Ayodeji Oginni. Maternal and fetal outcomes after introduction of magnesium sulphate for treatment of preeclampsia and eclampsia in selected secondary facilities: a low-cost intervention. Matern Child Health J 2013;17(7):1191–8.

[74] Boggess Kim A, Lieff Susi, Murtha Amy P, Moss Kevin, Beck James, Offenbacher Steven. Maternal periodontal disease is associated with an increased risk for preeclampsia. Obs Gynecol 2003;101(2):227–31.

[75] Le Quynh-Anh, Akhter Rahena, Coulton Kimberly Mathieu, Vo Ngoc Truong Nhu, Duong Le Thi Yen, Nong Hoang Viet, Yaacoub Albert, Condous George, Eberhard Joerg, Nanan Ralph. Periodontitis and preeclampsia in pregnancy: A systematic review and meta-analysis. Matern Child Health J 2022;26(12):2419–43.

[76] Bearblock Elizabeth, Aiken Catherine E, Burton Graham J. Air pollution and pre-eclampsia; associations and potential mechanisms. Placenta 2021;104:188–94.

[77] Bilenko Natalya, Ashin Michal, Friger Michael, Fischer Laura, Sergienko Ruslan, Sheiner Eyal. Traffic noise and ambient air pollution are risk factorsfor preeclampsia. J Clin Med 2022;11(15):4552.

[78] Wikström Anna-Karin, Stephansson Olof, Cnattingius Sven. Tobacco use during pregnancy and preeclampsia risk: effects of cigarette smoking and snuff. Hypertension 2010;55(5):1254–9.

[79] Sugimoto Yukihiko, Yamasaki Atsushi, Segi Eri, Tsuboi Kazuhito, Aze Yoshiya, Nishimura Tatsuya, Oida Hiroji, Yoshida Nobuaki, Tanaka Takashi, Katsuyama Masato, et al. Failure of parturition in mice lacking the prostaglandin f receptor. Science 1997;277(5326):681–3.

[80] Olson David M, Zaragoza DB, Shallow MC, Cook Jocelyn L, Mitchell BF, Grigsby P, Hirst J. Myometrial activation and preterm labour: evidence supporting a role for the prostaglandin f receptor—a review. Placenta 2003;24:S47–54.

[81] Hamshaw Isabel, Straube Anne, Stark Richard, Baxter Laura, Alam Mohammad T, Wever Walter J, Yin Jun, Yue Yong, Pinton Philippe, Sen Aritro, et al. PGF2α induces a pro-labour phenotypical switch in human myometrial cells that can be inhibited with PGF2α receptor antagonists. Front Pharmacol 2023;14:1285779.

[82] Hagberg Aric, Swart Pieter J, Schult Daniel A. Exploring network structure, dynamics, and function using networkx. Los Alamos National Laboratory (LANL), Los Alamos, NM (United States); 2008.

[83] visjs contributors. Vis-network: A dynamic, browser-based network visualization library. 2025, https://github.com/visjs/vis-network. (Accessed 12 September 2025).

[84] Torgano Francesco, Soto Gomez Mauricio, Zignani Matteo, Gliozzo Jessica, Cavalleri Emanuele, Mesiti Marco, Casiraghi Elena, Valentini Giorgio. RNA knowledge-graph analysis through homogeneous embedding methods. Bioinform Adv 2025;5(1):vbaf109.

[85] Romero Roberto, Yeo Lami, Chaemsaithong Piya, Chaiworapongsa Tinnakorn, Hassan Sonia S. Progesterone to prevent spontaneous preterm birth. Semin Fetal Neonatal Med 2014;19(1):15–26.

[86] Shah Nirav R, Bracken Michael B. A systematic review and meta-analysis of prospective studies on the association between maternal cigarette smoking and preterm delivery. Am J Obstet Gynecol 2000;182(2):465–72.

[87] Liu Buyun, Xu Guifeng, Sun Yangbo, Qiu Xiu, Ryckman Kelli K, Yu Yongfu, Snetselaar Linda G, Bao Wei. Maternal cigarette smoking before and during pregnancy and the risk of preterm birth: A dose–response analysis of 25 million mother–infant pairs. PLoS Med 2020;17(8):e1003158.

[88] Lin Jing, Jin Hua, Chen Lei. Associations between insulin resistance and adverse pregnancy outcomes in women with gestational diabetes mellitus: a retrospective study. BMC Pregnancy Childbirth 2021;21(1):526.

[89] Choudhury Abbas Alam, Rajeswari V Devi. Polycystic ovary syndrome (PCOS) increases the risk of subsequent gestational diabetes mellitus (GDM): A novel therapeutic perspective. Life Sci 2022;310:121069.

[90] Hedderson Monique M, Darbinian Jeanne, Havel Peter J, Quesenberry Charles P, Sridhar Sneha, Ehrlich Samantha, Ferrara Assiamira. Low prepregnancy adiponectin concentrations are associated with a marked increase in risk for development of gestational diabetes mellitus. Diabetes Care 2013;36(12):3930–7.

[91] McCarthy John F, Misra Dhirendra N, Roberts James M. Maternal plasma leptin is increased in preeclampsia and positively correlates with fetal cord concentration. Am J Obstet Gynecol 1999;180(3):731–6.

[92] El Shahat Amal Mohamed, Ahmed Abeer Bahaa, Ahmed Magdy Refaat, Mohamed Heba Saber. Maternal serum leptin as a marker of preeclampsia. Arch Gynecol Obstet 2013;288(6):1317–22.

[93] Doyle Lex W, Crowther Caroline A, Middleton Philippa, Marret Stephane, Rouse Dwight. Magnesium sulphate for women at risk of preterm birth for neuroprotection of the fetus. Cochrane Database Syst Rev 2009;(1).

[94] Crowther Caroline A, Brown Julie, McKinlay Christopher JD, Middleton Philippa, Pregnancy Cochrane, Group Childbirth. Magnesium sulphate for preventing preterm birth in threatened preterm labour. Cochrane Database Syst Rev 1996;2014(8).

[95] Lin Qimei, Cao Jiasong, Yu Jing, Zhu Yu, Shen Yongmei, Wang Shuqi, Wang Yixin, Liu Zhen, Chang Ying. YAP-mediated trophoblast dysfunction: the common pathway underlying pregnancy complications. Cell Commun Signal 2023;21(1):353.

[96] Gyamfi-Bannerman Cynthia, Pandita Ambika, Miller Eliza C, Boehme Amelia K, Wright Jason D, Siddiq Zainab, D'Alton Mary E, Friedman Alexander M. Preeclampsia outcomes at delivery and race. J Maternal-Fetal Neonatal Med 2020;33(21):3619–26.

[97] Borders Ann EB, Wolfe K, Qadir S, Kim Kwang-Youn, Holl Jane, Grobman William. Racial/ethnic differences in self-reported and biologic measures of chronic stress in pregnancy. J Perinatol 2015;35(8):580–4.

[98] Graves Cornelia R, Firoz Tabassum, Smith Skylar N, Hernandez Natalie, Haley Shaconna, Smith Kim, D'Oria Robyn, Celi Ann C. Addressing racial disparities in the hypertensive disorders in pregnancy: a plan for action from the preeclampsia foundation's racial disparities task force. J Racial Ethn Health Disparities 2024;1–10.