Improving Risk Assessment of Miscarriage during Pregnancy with Knowledge Graph Embeddings

Hegler C. Tissot · Lucas A. Pedebos

Abstract Miscarriages are the most common type of pregnancy loss, mostly occurring in the first 12 weeks of pregnancy. Pregnancy risk assessment aims to quantify evidence to reduce such maternal morbidities, and personalized decision support systems are the cornerstone of high-quality, patient-centered care to improve diagnosis, treatment selection, and risk assessment. However, data sparsity and the increasing number of patient-level observations require more effective forms of representing clinical knowledge to encode known information that enables performing inference and reasoning. Whereas knowledge embedding representation has been widely explored in the open domain data, there are few efforts for its application in the clinical domain. In this study, we contrast differences among multiple embedding strategies, and we demonstrate how these methods can assist in performing risk assessment of miscarriage before and during pregnancy. Our experiments show that simple knowledge embedding approaches that utilize domain-specific metadata perform better than complex embedding strategies, although both can improve results comparatively to a population probabilistic baseline in both AUPRC, F1-score, and a proposed normalized version of these evaluation metrics that better reflects accuracy for unbalanced datasets. Finally, embedding approaches provide evidence about each individual, supporting explainability for its model predictions in such a way that humans understand.

H. Tissot
University College London, UK
C3SL, Universidade Federal do Paraná (UFPR), Brazil
University of Pennsylvania, USA
E-mail: h.tissot@ucl.ac.uk; hegler@seas.upenn.edu

L. Pedebos
Secretaria Municipal de Saúde de Florianópolis, Brazil
Universidade Federal de Santa Catarina (UFSC), Brazil

1 Introduction

Over 200 million women become pregnant each year worldwide, and more than one fourth of these are estimated to end in pregnancy loss, even in developed regions as shown by some estimates [13, 2]. A high-risk pregnancy threatens the health (or life) of both mother and fetus, requiring specialized care. Some pregnancies become high-risk as they progress, whereas some women are at increased risk for complications even before they get pregnant for a diverse set of reasons, including (a) existing health conditions (e.g. high blood pressure, diabetes, or being HIV-positive), (b) overweight and obesity, that can lead to other complications, such as high blood pressure, preeclampsia, and gestational diabetes, (c) multiple births, and (d) old maternal age, mostly due chromosomal abnormalities [16, 23].

Pregnancy risk assessment aims to quantify evidence on risks and risk–outcome associations to reduce infant and maternal morbidity by influencing maternal behaviors before, during, and immediately after pregnancy. Maternal morbidities include any condition that is attributed to or aggravated by pregnancy and childbirth which has a negative impact on the woman's well-being and/or functioning, such as hemorrhage, sepsis, eclampsia, complications of obstructed labor, and miscarriage [8].

A study investigating 14,000 women in China [46] points out postpartum hemorrhage, hypertension dur-

ing pregnancy, diabetes, and anemia as the top observed pregnancy-related complications within that cohort, mainly correlated with delivery times, gestational weeks, and informal pregnancy examinations, whereas age at pregnancy, obesity, pre-existing medical conditions, number of pregnancies, and education were less significant, though still correlated. However, these cause-patterns can vary due to demographic factors, such as when comparing high- and low-income countries [20], or when comparing maternal race and ethnicity [36].

Miscarriages are the most common type of pregnancy loss and are both physically and emotionally painful. Mostly occurring in the first 12 weeks of pregnancy, they go usually unnoticed most of the time due to the delay in the perception of pregnancy by most women [3,35]. However, there are known factors of different natures related to the increase in the occurrence of spontaneous abortions, such as abnormal fetus development, use of drugs that interfere with pregnancy, diabetes mellitus, and previous kidney problems [7].

Although the number of high-quality health studies has made it possible to develop several evidence-based guidelines throughout the last years, there is still a lack of refinement in the use of case-oriented data in everyday clinical practice. Personalized medicine has the potential of switching healthcare standards from common guidelines to solid computational models based on data obtained for an individual patient to improve diagnosis, treatment selection, and health system efficiency. This is a new frontier health professionals and decision makers are facing, in which computational methods based on large sets of patient-centric data could become a new standard for clinical evaluation and prediction [33,34].

In [45], for example, authors propose a disease risk prediction model that uses historical medical diagnoses to compound a comorbidity network to both generate a risk prediction and provide explainable rules to support clinicians to understand the resulting risk pathway. Their assumption is that diseases can progress and co-occur according to the latent relationships of underlying mechanisms and that historical medical disorders can largely affect the future onset of certain diseases. Although claiming to achieve high accurate results, they do not provide supporting evidence about the dataset in terms of balance of positive and negative cases, which makes the reported AUC scores unreliable. Moreover, they do not describe whether each diagnosis was associated with a corresponding temporal reference, as the order of multiple events along the timeline can be a key factor when assessing patient risk. Lastly, additional relevant socio-demographic variables were not considered, weakening the resulting model on not being able

to assess their inherent risks, such as those related to aging or other clinical factors (e.g., BMI).

Due to the large number of patient-level observations and data sparsity, a new machine learning paradigm based on 'knowledge graphs' (KG) has recently emerged, which considers flexible rather than fixed sets of features used to describe each individual [6]. The flexible schema used to store information in the form of relationships (edges) between entities (nodes) enables the interactions of multiple clinical factors to be analyzed more effectively, moving from population-based approaches to patient-centric predictions [33,34].

Many methods for learning knowledge embedding representation (KER) have been developed. KER operates by (i) learning a low-vector representation of KG constituents (nodes and edges) that preserves the graph inherent structure and the semantics of different types of associations between entities, and (ii) exploiting low-rank latent structure in the data to encode known information that enables inference. Hence, KER models fit for performing analytical and predictive tasks that verify and infer multiple types of interactions between entities, as when representing complex patient-centric clinical knowledge [28].

In previous KER studies focused on open-domain tasks using more sparse data, although benchmark evaluation has steadily improved, approaches are complex, and their latent structure still remains unexplained [42]. Recent studies have shown that more simple approaches that incorporate ontological constraints favoring embedding quality are able to speed up the training process and achieve better performance when evaluated in similar benchmark tasks for denser domain-specific clinical datasets [39,10]. In this work, we evaluate whether KER approaches could assist on clinical risk assessment, and we test the ability of embedding approaches on both capturing the semantic correlations from multi-relational categorized data and performing inference. We demonstrate that even simple embedding approaches that utilize domain-specific metadata can improve the risk evaluation both before and during the earlier pregnancy stages.

2 Motivation

Clinical risk assessment systems support researchers and decision makers to understand the phenomena involving the target audience helping to build and review public policies focused on this group. Healthcare professionals can also be benefited when integrating Electronic Health Record (EHR) systems as an element of clinical decision support [44,32]. In both cases, a more assertive and less generalized risk score rather brings
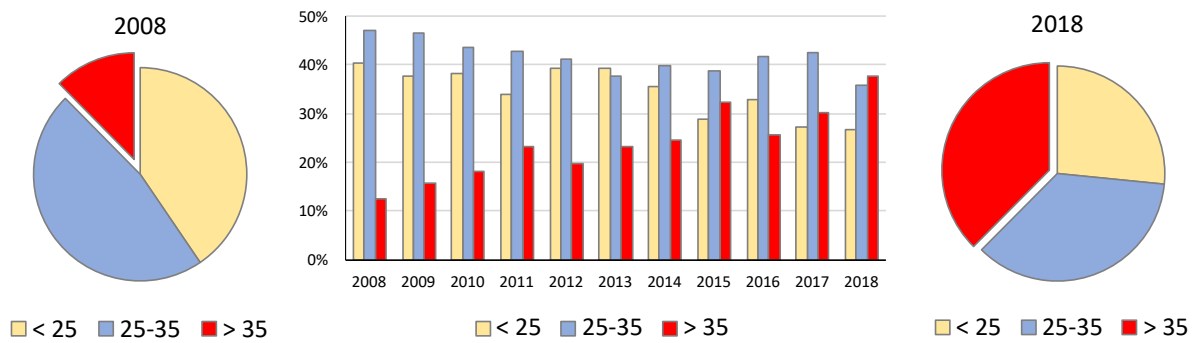
Fig. 1 2008-2018 distribution of miscarriages per maternal age (data report extracted from the *InfoSaude* system): (a) in 2008, miscarriages in women over 35-year-old represented only 12% of the cases; (b) over the years, the proportion of miscarriages in older women consistently increases; (c) in 2018, miscarriages in women over 35-year-old represented almost 40% of the cases.

the possibility of better using existing resources with a positive impact on the individuals' health. Finally, other benefits include the reduction of costs in clinical measures (e.g., screening and changing the drug of choice for cheaper ones) and clinical-administrative tasks (e.g., requesting routine tests with defined periodicity or repeat prescription for compensated chronic patients) [37].

Although global risk assessment can provide an overall picture for a given comorbidity, some populations can be under-represented and geo-related factors disregarded. Thus, the use of regional rather than national data is justified. Locally led research can support the evidence to understand the temporal changes in risk-attributable components (e.g., population growth, changes in population age, and changes in exposure to behavioral metabolic) when primary health care systems provide patient-centric individualized historical series coupled with a controlled and high-quality stable scenario that supports performing data analysis and designing more representative comparable risk assessment models [12].

In 2018, the Brazilian Ministry of Health elected the municipality of Florianópolis as the best primary health care capital within the National Program for Improving Access and Quality in Primary Care (PMAQ) [27], ranked it among the top-3 cities with the best Performance Index of the Unified Health System (ID-SUS). Moreover, Florianópolis has been using electronic medical records over the last 20 years, from which over 80% of the health network data being stored in digital format since 2008. *InfoSaude* [4, 40] is an Electronic Health Record (EHR) system created to manage and track medical records used to meet the needs of Florianópolis' 75 public health centers, integrating patient EHRs with multiple information structures, such as distinct types
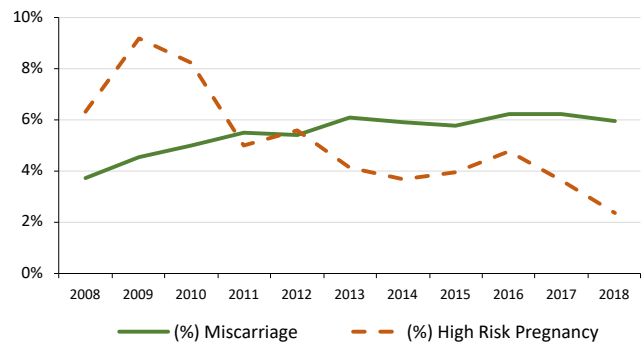

Fig. 2 Proportion of miscarriages vs. high-risk pregnancies contrasted over the period 2008-2018 (data report extracted from the *InfoSaude* system).

of care, pregnancies, procedures performed on each patient, applied vaccines and drug prescriptions.

Analysis of miscarriage reports from *InfoSaude* (Fig. 1) shows that the distribution of miscarriages regarding maternal age has been changing over the last decade. The proportion of cases in women under 25 decreased from 40% to 27% between 2008 and 2018, whereas it increased from 12% to 38% in the same period for women over 35. The former reflects some of the public health policies associated with specific early-aged pregnancy programs implemented over the last 15 years. The latter, however, evinces the odds in healthcare efforts and priorities. Finally, pregnancy losses in [25-35]-year-old women tend to remain proportionally constant between 2008-2018, with a slight increase in the period that coincides with spread of Zika virus in Brazil (2015-2016), with an estimated birth reduction of 7.78% response to Zika virus-associated microcephaly rate [15].

Changes in the patterns of socioeconomic factors can also play important role on significantly increasing the number of pregnancies in older women [1]. Moreover, pregnancy after age 35 that makes certain com-

plications more likely to directly reflect an expecting increasing number of high-risk pregnancies. However, within the *InfoSaude* cohort, there are several factors contributing to reduce the proportion of high-risk pregnancies in the last decade, which include (a) improvements in the quality of primary care services and healthcare programs, (b) use of well-defined health protocols leading to better control of chronic illnesses, and (c) adoption of smoking reduction programs. The latter, for example, allowed a reduction in the proportion of female smokers from 15.8% to 8.6% in the period between 2006-2016, as reported in a study that estimated sociodemographic frequency and distribution of risk and protective factors for chronic diseases in the capitals of the 26 Brazilian states and the Federal District in 2016 [29]. Indeed, Fig. 2 shows that the percentage of pregnancies flagged as 'high-risk' has decreased from 6.5% to less than 3.0% within a 10-year period. However, the corresponding number of miscarriage cases has increased from 3.7% up to 6.2% in the same period, which raises an alert on how pregnancy risk assessment protocols have been effectively applied.

3 Materials & Methods

3.1 Dataset

In this work, we combined structured and unstructured data extracted from the *InfoSaude* system containing 24,877 pregnancies occurring in the period from 2010 to 2015 to evaluate how KER approaches could better assist on performing risk assessment of miscarriage during pregnancy.[1]

Structured data covering most of the demographic and clinical history before and during pregnancy was coupled with additional unstructured information relevant to this use-case, the latter extracted from short notes about history of allergies, infections, and other clinical conditions. Such notes are written by physicians usually limited to 50-100 characters. Very simple regular expressions were able to extract the most common facts as well as the language signals for eventual negations. Finally, a hybrid phonetic similarity algorithm [18,40] was used to find and merge misspellings, positively identified and manually checked for about 1.5% of the mentions in each considered set of substances and infections.

Infections during pregnancy are a frequent cause of morbidity and mortality among mothers, fetuses, and

neonates, leading to miscarriages, preterm births or fetal deaths, and fetal malformations and growth restrictions. Infections can be asymptomatic or manifest symptoms like those in non-pregnant individuals, and their risk of causing pregnancy outcomes depends on factors such as (a) prevalence of infections in the population, (b) socioeconomic and cultural factors, and (c) health and healthcare-seeking behavior [17].

For each pregnancy, we collected a set of 19 categorical and 4 numerical features, subsequently mapped as relations in the KG. Numerical features were normalized accordingly: a) age in ranges of 5 years, b) weight in ranges of 5 kilograms, c) height in ranges of 5 centimeters, and d) BMI rounded to an integer value. However, we found there is not a systematic way of collecting information for several features considered in this study, for example, history of infections within the analyzed population, from which only 2.5% of the pregnancies have corresponding entries recorded in their EHR. Coverage (%) below is given to each feature in which data is not available for all cases. Features can have low coverage because of multiple factors, including non-required fields in the system (e.g., occupation), low probability of a true positive case (e.g., drug allergies and infections), and when pregnancy follow-up starts later, few weeks after the last menstrual period (LMP), and LMP-corresponding features are not measured (e.g., weight, height, and BMI on LMP).

– Demographic data: year when pregnancy started, age in years on the corresponding LMP's first day, marital status, ethnicity, education level, neighborhood, profession (36.7%), occupation (7.1%), country of birth (1.8% not from Brazil), and assigned local health office.
– Clinical history: weight (39.5%), height (34.3%) and BMI (25.6%) on LMP, the most common known infections (2.5%) ∈ {HIV, Bronchitis, Candidiasis, Hepatitis, Pyelonephritis, Syphilis, Toxoplasmosis, Urinary, Vaginitis, Vaginosis}, the most common known drug allergies (5.1%) ∈ {AAS, Iodine, Acetaminophen, Amoxicillin, Ampicillin, Benzetacil, Buscopan, Diclofenac, Dipyrone, Metoclopramide, Nimesulide, Penicillin, Plasil, Sulfa}, and other known clinical conditions (12.2%) ∈ {alcohol abuse, smoker, anemia, hypertension}.
– Prescriptions: each medication is identified by the generic name of the substance used – a total of 170 substances are being considered in this dataset.
– Diagnoses: each diagnosis is given by the ICD-10 code coupled with the corresponding appointment's service group (an administrative concept used within the *InfoSaude* system) and the physician's medical specialty.

---

[1] Data controllers of the *InfoSaude* system have granted us permission to use and perform analysis on a de-identified version of this dataset, and no time limit has been set for data usage.

Table 1 Statistics of training (LMP from 2010 to 2014) and test sets (LMP occurring in 2015): total number of pregnancies in each set and number of triples per relation type in the KG occurring before (pre) and after (post) LMP.

Stats	Training Set	Test set
Years	2010-2014	2015
Pregnancies	20,201	4,676
Triples		
(pre LMP)		
Demographics	166,436	37,266
Clinical history	63,745	14,376
Prescriptions	103,010	20,882
Diagnoses	143,667	31,669
Procedures	54,520	13,385
(post LMP)		
Weeks 01-08	107,870	25,380
Weeks 09-16	197,831	45,403
Weeks 17-24	191,133	45,226

– Procedures: each procedure is coupled with the corresponding nurse's or physician's medical specialty.

The time window we collected data regarding prescriptions, diagnoses, and procedures is given by: (a) before pregnancy, we considered a one-year period split into the corresponding quarters (Q1-Q4) immediately before LMP, and (b) during pregnancy, we used the first 24-week pregnancy period split into three periods of eight weeks each (B1, B2, and B3). Each split adds a variation in the name of each relation to incorporate a temporal context, so that each relation (e.g., `Prescription`) becomes a set of seven distinct relations (chronologically, `Prescription_Q4` to `Prescription_Q1` regarding the one-year period before LMP, and then `Prescription_B1` to `Prescription_B3` regarding the 24-week period after LMP). Therefore, if the same drug is prescribed twice, for example, about one year before pregnancy and in the first weeks of pregnancy, it will appear in the KG represented by two distinct relations, `Prescription_Q4` and `Prescription_B1` respectively. The number of triples in each pre- and post-LMP period is given in Table 1, and Fig. 3 shows an example on how a pregnancy is represented as a set of triples (subject, relation, object) in the resulting dataset.

Training and test sets were evenly and cumulatively split into: (a) LMP includes demographic data, clinical history and the one-year history of prescriptions, diagnoses and procedures preceding the first day of the LMP for each confirmed pregnancy; (b) B1 adds prescriptions, diagnoses and procedures extracted from the first 8 weeks following LMP; (c) B2 adds prescriptions, diagnoses and procedures extracted from weeks 9-16 following LMP; and (d) B3 adds prescriptions, diagnoses and procedures extracted from weeks 17-24 following LMP. This setup allowed us to assess risk in four different points in time, starting from the beginning of preg-

```
// Demographics
triple( Pregnancy:e77d, Age, AgeYears:25 )
triple( Pregnancy:e77d, Ethnicity, Ethnicity:4 )
triple( Pregnancy:e77d, MaritalStatus, MStatus:Married )
triple( Pregnancy:e77d, MaxEducation, Educ:Graduated )

// Medical History
triple( Pregnancy:e77d, Infection, Infection:Urinary )
triple( Pregnancy:e77d, Allergy, Allergy:Dipyrone )

// Pre-Pregnancy Diagnoses up to 3 months before LMP
triple( Pregnancy:e77d, ICD_Q1, ICD:L50 )
triple( Pregnancy:e77d, ICDSpec_Q1, Spec:Physician )

// Pre-Pregnancy Procedures up to 3 months before LMP
triple( Pregnancy:e77d, Procedure_Q1, Proc:9391910064 )

// Pre-Pregnancy Prescriptions up to 3 months before LMP
triple( Pregnancy:e77d, Prescr_Q1, Med:Loratadina )

// Pre-Pregnancy Prescriptions 3-6 months before LMP
triple( Pregnancy:e77d, Prescr_Q2, Med:Prednisone )

// Post-Pregnancy Prescriptions weeks 1-8 after LMP
triple( Pregnancy:e77d, Prescr_B1, Med:FolicAcid )

// Post-Pregnancy Prescriptions weeks 9-16 after LMP
triple( Pregnancy:e77d, Prescr_B2, Med:Dimenidrinate )
```

Fig. 3 Example of pregnancy represented by a set of triples extracted from the KG dataset – pregnancy ID was synthetically generated, and triples in this example were randomly selected from multiple patients to avoid identification.

nancy, and going through the first 6 gestational months. Although labels that flag miscarriages are not part of the datasets, we found data signals in the KG (e.g., ICD-10 diagnosis codes corresponding to miscarriage) that could potentially influence scores and bias the results favoring positive labeled cases in later pregnancy stages. Thus, pregnancies in which the miscarriage occurred in any of the three periods of 8 weeks after LMP were removed from all subsequent test sets since they happened.

3.2 Knowledge Embedding Representation

Knowledge Graphs (KGs) are a widely used representation for multi-relational data, comprising entities (nodes) and relations (edges) that provide a flexible structured schema adapted for both open- and domain-specific knowledge bases. Embedding representation methods can (a) learn and operate on the latent feature representation of the KG constituents and on their semantic relatedness, (b) efficiently handle data sparsity and inconsistency, such as missing or multiple inconsistent values for 'one-to-one' relations, and (c) enhance subsequent machine learning applications.

TransE [5] is a well-known translational approach that uses simple assumptions to achieve considerably accurate and scalable results, proved to be an effective and efficient embedding model [22]. Several enhanced

models (e.g., TransH [43], TransR [25], Rescal [31], Complex [41], and HolE [30]) were proposed to address some of the TransE's supposed flaws on representing 'many-to-many' relations. These methods attempt to couple TransE with more complex relation-specific representation of different data cardinality (e.g., projection matrices) to improve performance on the link prediction benchmark task (also known as knowledge graph completion). However, we show in previous work [10,11] that these models do not necessarily provide a good representation in terms of embedding quality, in which similar entities are expected to be found in small clusters along the resulting embedding space. Indeed, embedding methods in general tend to report weak performance [42,19].

Besides, we observed other well-known state-of-the-art models in open-domain benchmark tasks, such as Rescal, HolE and Complex, can take up to 30-50 times more CPU resource to perform training, whereas is still unclear the ability of these models to properly embed domain-specific knowledge representation. Experimental results testing these models with clinical datasets showed that, even after extending the training process over the usual 1000 learning epochs, the standard embedding evaluation protocol still shows very poor accuracy for dense clinical KGs. Moreover, traditional embedding systems are designed to tackle the evaluation task instead of favoring embedding quality, and the resulting embedding representation tends to cluster entities by their use as head or tail in each relation instead of their semantic relatedness. As a side-effect, groups of entities (similar type) are formed in such a way that makes the link prediction task hard, even when types are not explicitly given as part of the KG metadata.

Given the known flaws of translational models and their limitations on using denser domain-specific datasets, HEXTRATO [39] emerged a translational embedding approach that couples TransE with a set of ontology-based constraints to learn representations for multi-relational categorized data, originally designed to embed biomedical- and clinical-related datasets. It improves the translational embedding from TransE and other enhanced models, achieving great performance, even in very low k-dimensional spaces ($k < 100$), without necessarily adding complex representation structures within the model training process, such as those used in TransE-based enhanced models.

In TransE, head and tail entities (h and t) and relations (r) are represented by translation vectors $\mathbf{h}, \mathbf{t}, \mathbf{r} \in \mathbb{R}^k$, chosen so that every relation r is regarded as a translation between h and t in the embedding space. The pair of embedded entities in a triple (h, r, t) can be approximately connected by $\mathbf{r}$ with low error (Equa-

tion 1). TransE does not take into account any metadata as part of the knowledge graph, whereas HEXTRATO uses types to identify each entity. The latter proved useful to accelerate and improve training and representation [39].

$$\mathbf{h} + \mathbf{r} \approx \mathbf{t} \qquad (1)$$

For categorized datasets, given a training set S of categorized triples $(c_h{:}h,\, r,\, c_t{:}t)$, embedding vectors for entities and relations are learned, so that each categorized entity $c{:}e$ is represented by an embedding vector $\mathbf{e_c} \in \mathbb{R}^K$, and each relation r is represented by an embedding vector $\mathbf{r} \in \mathbb{R}^K$. A score function f_r (Equation 2) represents a L2-norm dissimilarity, such that the score $f_r(h_{c_h}, t_{c_t})$ of a plausible typed triple $(c_h{:}h,r,c_t{:}t)$ is smaller than the score $f_r(h'_{c_h}, t'_{c_t})$ of an implausible typed triple $(c_h{:}h',r,c_t{:}t')$. Then, the optimal KER model is learned by minimizing a margin-based (γ) loss function $\mathcal{L}$ (Equation 3) adapted from TransE, where γ is the margin parameter, $\mathcal{S}$ is the set of correct triples, and $\mathcal{S}'$ is the set of incorrect triples $(c_h{:}h',r,c_t{:}t) \cup (c_h{:}h,r,c_t{:}t')$.

$$f_r(h_{c_h}, t_{c_t}) = \|h_{c_h} + r - t_{c_t}\|_{l_2} \qquad (2)$$

$$\mathcal{L} = \sum_{\substack{(c_h{:}h,r,c_t{:}t) \in \mathcal{S} \\ (c_h{:}h',r,c_t{:}t') \in \mathcal{S}'}} [\gamma + f_r(h_{c_h}, t_{c_t}) - f_r(h'_{c_h}, t'_{c_t})]_+ \qquad (3)$$

Finally, in both TransE and KRAL, we mimic enhanced models by introducing a relation-based projection matrix $M_r \in \mathbb{R}^{K \times K}$ in the scoring function used during training (Equation 4).

$$f'_r(h_{c_h}, t_{c_t}) = \|M_r \times h_{c_h} + r - t_{c_t}\|_{l_2} \qquad (4)$$

Link prediction (LP) is a traditional evaluation protocol also used during training when learning knowledge embedding representation. However, in previous work [10], we evaluated whether LP accurately reflects the quality of the resulting embeddings for multi-relational data, showing that more complex embedding representation approaches used to improve KG completion do not reflect into embedding quality. Thus, we extended HEXTRATO with additional features to mimic enhanced TransE models that use projection matrices, but still keeping the idea of using set of independent type-associated hyperspaces in order to project each entity belonging to the same type. We developed KRAL,[2] an embedding

---

[2] https://github.com/hextrato/KER

framework that allows implementing and contrasting multiple embedding representation methods, and we used this framework to mimic and simulate what most of the methods enhancing TransE use as embedding strategy, coupling the vectors resulting representation of each relation from TransE with projection matrices that aim to improve accuracy in traditional benchmark datasets. The four approaches considered along our experiments are described as follows:

- TransE: the original TransE approach using a vector representation for each relation;
- TransE+: a projection matrix is added to the embedding representation of each relation, mimicking some of the TransE enhanced approaches (e.g., TransH and TransR);
- KRAL: this is the original HEXTRATO approach, in which independent hyperspaces are used for each type of entity, and similarly to TransE, it uses a vector representation for each relation;
- KRAL+: like in TransE+, a projection matrix is added to the embedding representation of each relation.

3.3 Risk Assessment

In order to generate a risk score for each pregnancy we look at the embedding neighborhood of each pregnancy. Within the resulting 64-dimensional embedding hyperspace with radius 1.0, we use multiple maximum L_2-norm radiuses ranging from 0.125 to 1.5 to calculate the percentage of neighbors labeled as positive to miscarriage (32 distinct values in total). Then, we compare the resulting F1-score from each neighborhood radius.

Very small radiuses can eventually find no neighbor entities, whereas very high radiuses will consider too many entities, so that the resulting ratio of positive cases would approximate the overall percentage of miscarriages in the entire dataset. Within our experiments, we found the risk assessment reached best F1 scores when using $radius = 0.321025$ during tuning.

We used three baselines for miscarriage risk assessment. Firstly, a 'Random' risk score representation is created to serve as primary baseline. Then, a population-based approach uses pregnancy feature-value pairs to provide a feature-based probabilistic score, presented as 'Prob(F)' within our results. Finally, a logistic regression (LogReg) approach.

Prob(F) takes into account all $P(f, v)$ supported by a minimum number of cases. $P(f, v)$ corresponds to the probability of the target label (a miscarriage in this case) happening when a feature $f \in F$ has a value $v \in V(F)$. For example, $P(AgeYears, 45+)$ (0.2302) is the

probability of a miscarriage happening in a patient over 45 years-old, and $P(PrescriptionPrevQ1, Fluoxetine)$ is the probability of a miscarriage happening for a patient prescribed with Fluoxetine in the quarter preceding LMP (0.1135). We only consider features supported by a minimum number of 86 cases, which is the minimum sample size needed to reach a confidence level of 95% when resulting scores are within ±5% of the measured/surveyed value, for a training population of 20,201 cases and the proportion of cases labeled as miscarriages is 5.91%. For each pregnancy, the probabilities for all known pairs (f, v) fitting the previous condition are averaged to provide a final risk assessment score.

All choices to validate the test set (2015's pregnancies), including the radius for finding similar pregnancy cases (cluster) and the risk assessment threshold from resulting scores in the range [0, 1], were made based on a tuning set corresponding to 2014's pregnancies, except for LogReg, in which we used the threshold 0.5 when determining whether a resulting risk score correspond to a positive or negative miscarriage.

3.4 Evaluation

High class imbalanced data and skewed data distributions are naturally observed in many real-world scenarios, such as in medical diagnosis-related tasks, in which most of the cases represent healthy patients as the negative class. Imbalanced data poses further challenges related to bias towards the majority class, even ignoring the minority target class in extreme cases, typically (a) under-classifying the minority group, and (b) misleading result with high scores that incorrectly indicate good performance [21].

Receiver operating characteristics (ROC) curve is a usual assessment to contrast true vs. false positive rates. However, ROC curves can exhibit unduly optimistic results, as the false positive rate (FPR) is less sensitive to changes in false positive (FP) as longer the size of the negative class is predominant [14]. Precision-Recall curves (AUPRC) are recommended be used for imbalanced data instead [14].

In our dataset, the ratio between negative and positive cases is ≈ 16.91 (5.91% of positive cases), which means there are almost 17 negative cases for each positive one, which increases the probability of more false positive cases happening along the evaluation, worsening the results due to the unbalanced distribution of the target label. Thus, in addition to presenting the original AUPRC and F1 scores, we also present weight-adjusted scores relying on a normalized version of false positive (FP) cases, which better approximates the scores to

a balanced-similar scenario, potentially facilitating further analysis and comparison of embedding approaches on multiple datasets. F_{norm}-score uses FP_{norm} instead of FP:

$$FP_{norm} = \frac{FP_{test}}{(TN_{train}/TP_{train})} \quad (5)$$

where FP_{test} is the number of false positive cases in the test set, TN_{train} is the number of true negative cases in the training set, and TP_{train} is the number of true positive cases in the training set. FP_{norm} is used as a replacement to the original FP when calculating $AUPRC_{norm}$ and F_{norm} scores.

4 Results

The *InfoSaude* system reports 270 cases of miscarriage out of 4,676 pregnancies in which LMP occurred in 2015 (5.77%), corresponding to the test set in our experiments. Our method attempts at embedding a pregnancy knowledge graph to find clusters of similar patients that can support the decision of establishing a high-risk pregnancy.

In table 2 we present AUPRC, F1 and F_{norm} scores for each approach evaluated and Fig. 4 compares all approaches regarding AUPRC in each pregnancy phase, starting at the beginning of pregnancy (LMP), followed by each period of 8 weeks after LMP (B1–B3).

The random approach performs as expected in a binary classification task, with F_{norm} scores close to 0.5 along all pregnancy phases. Then, Prob(F) and LogReg approaches are given as efficient baseline candidate methods for risk assessment. Prob(F) combines the probabilities of miscarriages (target labels) occurring in each pair (f, v) of features f and their corresponding values v, performing F_{norm} scores in the range of $\approx [0.67, 0.69]$. Logistic regression performs better than TransE-like approaches in late pregnancy stages. Finally, translational embedding approaches can preserve the graph inherent structure and the semantics of different types of associations between entities, and even though F_{norm} scores improvement is relatively low regarding Prob(F) and LogReg, embedding methods are able to support analytical and predictive tasks that require inferring multiple types of interactions between entities that represent complex patient-centric clinical data. Embedding approaches that deal with non categorized open-domain data (TransE-like approaches) use a single hyperspace to accommodate all entities, whereas KRAL takes the advantage of multiple type-based independent hyperspaces to optimize the spread of entity vectors and the loss function during training, which makes it performs slightly better than other embedding
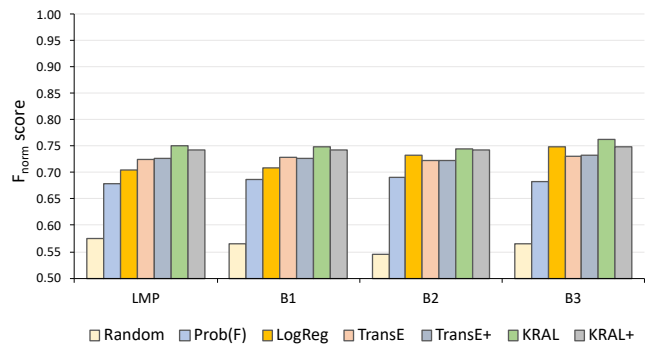

Fig. 4 Overall results: Weighted $F1_{norm}$ scores for translational embedding and baseline models.

strategies, reaching F_{norm} scores ≈ 0.75. Visualization for the resulting KRAL embedding representation is available online.[3]

In addition to analyzing model performance along the evaluation task, we also compare differences between each resulting model regarding statistical significance. We used a dependent sample paired Sample T-Test to determine whether the mean difference between two sets of observations can be considered statically significant. Table 3 presents the resulting p-value and t when comparing two models, considering significance level (alpha) = 0.05, H_0 (A=B) is rejected when p-value < alpha, and $|t|$ accepted range $[-1.9605 : 1.9605]$.

$F1_{norm}$ scores clearly show the improvement in performing risk analysis when moving from Random to Prob(F), and from the latter to embedding methods. However, it is still unclear whether more complex translational methods can improve the representativeness of a resulting embedding representation, as long as previous works have been mostly focused on link prediction as an evaluation task instead of assessing the quality of the resulting embedding models [10].

In open-domain data, the addition of projection matrices to TransE (generally represented in our experiments by TransE+) has not been able to add any significant signal that could be used to improve the evaluation task, and there were no significant statistical differences observed when comparing these models to the original TransE. Similarly, the same happens for translational models designed to deal with domain-specific data. Actually, the addition of projection matrices made KRAL+ to perform slightly worse than the pure linear translational approach used by KRAL.

Results reinforce our assumption that relation-based projection matrices do not necessarily favor embedding

---

[3] `https://projector.tensorflow.org/?config=https://raw.githubusercontent.com/HeglerTissot/hextrato/master/KER-View/PRAS.json`

Table 2 AUPRC, F1 and F_{norm} scores on the test set resulting from three baseline (BL) methods and four KER variations.

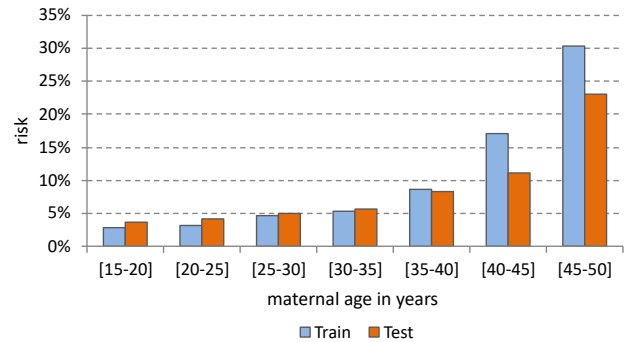| | | AUPRC | | | | F1-score | | | | F_{norm}-score | | | |
		LMP	Pregnancy Weeks			LMP	Pregnancy Weeks			LMP	Pregnancy Weeks		
			B1	B2	B3		B1	B2	B3		B1	B2	B3
BL	Random	0.032	0.025	0.032	0.033	0.115	0.112	0.109	0.112	0.575	0.565	0.545	0.564
	Prob(F)	0.038	0.042	0.052	0.087	0.133	0.138	0.165	0.122	0.679	0.687	0.692	0.683
	LogRegr	0.043	**0.049**	0.053	0.093	0.139	0.142	0.159	0.179	0.704	0.708	0.732	0.749
KER	TransE	0.051	0.029	0.046	0.057	0.125	0.124	0.121	0.126	0.724	0.729	0.724	0.731
	TransE+	0.015	0.042	0.030	0.087	0.122	0.148	0.135	0.189	0.726	0.726	0.723	0.733
	KRAL	**0.054**	0.043	**0.067**	**0.099**	**0.159**	**0.151**	**0.172**	**0.192**	**0.750**	**0.748**	**0.745**	**0.762**
	KRAL+	0.011	0.011	0.022	0.077	0.122	0.121	0.132	0.184	0.743	0.743	0.744	0.749

Table 3 Resulting significance tests between baselines and KER risk assessment models performed on the final scores over the test set.

Approaches		Split	p-value	t
Random	Prob(F)	LMP	-3.5527e-15	11.9903
		B1	0.0000e-00	33.3973
		B2	1.8874e-15	-15.1822
		B3	0.0076e-00	-2.6700
Prob(F)	TransE	LMP	3.1978e-10	-6.3025
		B1	0.0000e-00	-138.7982
		B2	0.0000e-00	-110.0519
		B3	0.0000e-00	-97.3523
TransE	TransE+	LMP	0.2691e+01	1.1052
		B1	0.3109e+01	1.0853
		B2	0.8873e+01	0.1418
		B3	0.1998e+01	-1.1492
KRAL	KRAL+	LMP	4.44089e-16	77.8029
		B1	1.11022e-16	-46.2310
		B2	1.11022e-15	12.3938
		B3	5.55112e-16	-34.9729
KRAL	TransE	LMP	-2.22045e-16	-27.0375
		B1	-1.05386e-15	-19.2891
		B2	-1.04290e-15	-16.2562
		B3	-2.22045e-16	-19.0259

Fig. 5 Overall probabilistic risk of miscarriage related to maternal age given by the training and test sets.
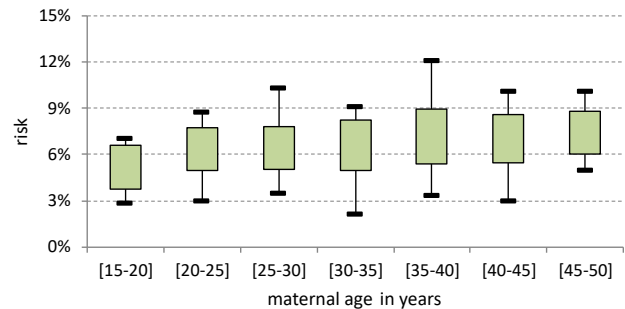
Fig. 6 Risk of miscarriage related to maternal age resulting from embedding neighbors' similarity – min, max, and avg±stdev miscarriage ratios considering up to 100 nearest training neighbors for each pregnancy in the test set.

quality. Non-similar entities can be placed very close to each other in opposite sides of any dimensional axis (opposite hyper-hemispheres). Therefore, the similar resulting vector representations blur their semantic dissimilarity when they are pre-projected to opposite directions by the relation matrix before having the relation vector added to their latent composition.

In TransE and other single-hyperspace embedding approaches derived from TransE, entities from one type can induce other types of entities to form unexpected clusters when competing for space (i.e. optimizing the loss function during training). KRAL uses independent hyperspaces to embed each type of entities, in such a way that all pregnancies will form clusters that represents either the disjointness or the similarity regarding all other pregnancies, independently of other entity types, such as medications or ICD-10 codes. We believe this is a useful embedding feature when trying to represent categorized multi-relational data in specific domains.

Despite the progress in understanding the causes of early pregnancy losses, there are mechanisms underlying maternal losses that remain unknown, mostly related to cellular or genetic meiotic chromosomal abnormalities, that can be translated into no observed correlation with sociodemographic factors and clinical history [26]. In this sense, KER provides the supporting evidence to assist finding common characteristics in a K-nearest neighborhood analysis of similar cases.

Fig. 5 shows the probabilistic risk distribution of miscarriage in each maternal age range for both training and test sets used in our experiments. Although pregnancy in older women is subject to higher risk, such age-related scores are only a generalization of the cases and do not provide any concrete evidence to support patient-centric decisions.

KER learns the latent feature representation of each pregnancy, making those with similar characteristics to be found in the nearby neighborhood within the hyperspace, based on their semantic relatedness. Fig. 6 shows a different scenario for the resulting risk assessment regarding each maternal age range. The average miscarriage ratios are similarly higher in all maternal ages from 35-years-old. However, pregnancy cases with the maximum ratios are observed in the range [35,40] instead of [45,50].

Unlike probabilistic risk assessment approaches, KER provides a distinct risk score for each pregnancy individually. Moreover, when compared to the most similar cases, the pairs of more relevant feature values can be analyzed to describe which characteristics are more relevant to the specific group of pregnancies semantically related to the assessed pregnancy. We randomly selected 15 pregnancy cases (all women over 30-years-old) and we manually analyzed the top-100 feature values for each assessed pregnancy cluster. Table 4 some of the most clinically relevant features for each case (manually selected by an experienced clinician). Pregnancy cases within the same maternal age can be described with very different features. For example, some pregnancy cases have higher assessed risk ratio and belong to a group of similar cases that share conditions known to be related to miscarriages (e.g., infections, smoking, and hypertension), whereas the use of Folic Acid is one of the features consistently describing those clusters corresponding to cases with lower risk ratio.

Finally, in Table 5, we show how the proposed method could improve the risk assessment of miscarriage and how it would reflect within the definition of a 'high-risk' pregnancy. Considering all cases of positive miscarriage in the test set, KRAL was able to increase the predicted number of pregnancies set as 'high-risk', overall, from 0.8% to 1.4%, and from 1.7% to 2.4% in the set of pregnant women over 30-years-old, meanwhile reducing the percentage of false negative high-risk pregnancies ('high-risk' = 'no').

5 Discussion

The findings in this study aim to identify a computational method that could support isolating the impact of each risk factor described and improve the safety of clinical decisions before and during the pregnancy comparatively to what current protocols can actually do. We claim that KER embedding methods have the ability of learn the low-dimensional latent feature representation of the KG constituents, operate on their semantic relatedness, and efficiently handle data sparsity and inconsistency issues. Subsequently, the result-

Table 4 Examples of relevant features obtained from the KRAL approach for a sample of 15 patients over 30-years-old. For example, 'Folic Acid' is shown as relevant for patents with lower risk (#3, #4, #8, #11, #13), whereas 'Smoker' is pointed out for some of the high-risk patients (#6, #10, #14, #15).

#	Age	Risk	Relevant features
1	30-35	9.1%	history of psychological and psychiatric treatment preceding LMP – diagnosis of 'Adjustment disorders' (ICD-10 F43.2) 9-12 months before pregnancy – infectious disease tested in the first 8 weeks of pregnancy
2	30-35	8.2%	use of Nimesulide and Ciprofloxacin in the quarter preceding pregnancy
3	30-35	3.4%	Tenoxicam and Amoxicillin prescribed between 6-12 months before pregnancy – using Folic Acid during pregnancy
4	30-35	2.1%	using Folic Acid
5	35-40	12.1%	history of psychological and psychiatric treatment preceding LMP – use of Fluoxetine in the year before pregnancy – Amoxicillin prescribed between 6-12 months before pregnancy
6	35-40	11.3%	smoker
7	35-40	4.0%	dental treatment during pregnancy
8	35-40	3.4%	smoker – using Folic Acid – OB-GYN visit in the first 8 weeks of pregnancy (high-risk considered)
9	40-45	10.1%	Amoxicillin prescribed between 6-12 months before pregnancy – psychological treatment in the quarter preceding LMP
10	40-45	9.8%	smoker – consistently using Amoxicillin in the 12 months preceding pregnancy – dental treatment during pregnancy – OB-GYN visit in the first 8 weeks of pregnancy (high-risk considered)
11	40-45	3.9%	smoker – using Folic Acid
12	40-45	3.0%	Amoxicillin prescribed between 6-12 months before pregnancy – using Folic Acid
13	45-50	10.1%	hypertension
14	45-50	8.1%	smoker – Amoxicillin prescribed between 6-12 months before pregnancy
15	45-50	6.1%	smoker – history of urinary infection – diagnosis of 'Infectious Gastroenteritis' (ICD-10 A09) 3-6 months before pregnancy – psychiatric treatment in the semester preceding LMP – using Folic Acid during pregnancy

ing embedding representation provides the evidence to support analyzing groups of semantically related pregnancies regarding their commonly observed features.

The availability and quality of maternal and newborn health care are the "litmus test" of health systems [8]. There is an increasing interest of clinicians and federal healthcare policy makers on modeling disease risk to identify and analyze potential events that

Table 5 Contrasting the performance of *InfoSaude* (observed) *vs.* KRAL regarding the overall pregnancy risk of miscarriage in the test set. *ISS=*InfoSaude* system

	All cases		Age > 30-years-old	
High-risk	**ISS**	**KRAL**	**ISS**	**KRAL**
Yes	0.81%	1.39%	1.70%	2.40%
No	4.96%	4.38%	6.70%	6.00%

may negatively impact patients, lessen subsequent patient health decay, make judgments about the tolerability of the risks while considering multiple influencing factors, and reduce preventable harm and associated healthcare costs [45].

Treatment recommendations in clinical practice guidelines are supported by the evidence collected from research studies that utilize populations with highly selective sociodemographic and comorbidity-related characteristics. However, when treating complicated patients, physicians need to determine the applicability of a study to their clinical population who do not wholly align with guideline recommendations [9]. A more accurate evidence-based risk correction, supported by policies and protocols defined from data analysis, allows the health system to focus on a multi-level care stratagem, instead of only using the existing binary decision approaches (low- and high-risk pregnancy).

Disease risk prediction models are preferable as longer the ability of constructing explainable rules that support understanding why and how the prediction was made. Revealing why the predicted result is made for a patient is an important condition in both heightening the patient's trust towards the model and help physicians to choose whether to trust the model, combining derived information from their domain knowledge in favor of accepting or rejecting the predicted result [45]. KER can efficiently provide evidence that supports a risk assessment scale, moving from the traditional binary decision to a continuous risk scale [0,1] to support a multi-level risk assessment.

Prenatal care before and during pregnancy significantly decreases the risk of complications that can lead both mother and developing fetus to death. Clinical decision support systems (CDSS) play an important role in all levels of the health system on assisting health providers in decision making that can improve the quality of prenatal care, identifying potential conditions that lead to high-risk pregnancy, and contributing to improve quality of health care [24]. We noted that some of the characteristics observed in the analyzed cases are not commonly highlighted with great weight in the literature, neither addressed as main elements in guidelines. The possibility of assigning more individualized values to these factors (unlike the typical 'high' or 'low' risk),

in addition to the patient-centered analysis, provides a more granular stratification of risks regarding miscarriages, which makes the interactions between physician and patient straighter and more accurate.

The uncertainties surrounding population studies, in many cases, left risk studies with only dichotomous classifications, such as 'high-risk' and 'low-risk', which is equally observed in the *InfoSaude* system. In this regard, our work contributes on providing a probabilistic risk scale in which multiple classes of risk can be designed over (e.g., 'none', 'low', 'medium', 'high', 'very high'). Moreover, it supports the task of identifying overall high-risk pregnancies in early stages rather than eventually being reactive to diagnosing specific pregnancy-related comorbidities.

Finally, decisions made solely based on algorithms that need to be constantly being trained from quality databases, without professionals knowing their functioning and limitations have been reported as elements that lead to the failure of the strategy or to error in clinical decision making [38,44]. Several strategies and guidelines for the use of risk assessment systems have been suggested based on experiences of greater or lesser success, as well as general models of what a CDSS can or cannot accomplish [47]. This seems to be the central point on maximizing the benefits and minimizing the implementation risks. In addition to defining the scope and guaranteeing the quality of the data used, the implementation strategy, with simple, clear and widely discussed elements with clinical professionals, seems to contribute to (a) better accept the system's suggestions, (b) know the limitations of the algorithms and (c) make rational use of the suggestions but adding to their clinical knowledge for decision making, instead of understanding the CDSS guidelines as a confrontation with their own knowledge, the opposites of these elements being some of the main risks involved.

6 Conclusions

Health problems during pregnancy can affect mother's health even when they were healthy before getting pregnant, and such complications may make it a high-risk pregnancy, caused by factors ranging from population demographics to clinical history. In this study, we show how knowledge embedding approaches can support risk assessment of miscarriage during pregnancy. We tested the ability of embedding approaches on capturing the semantic correlations from multi-relational categorized pregnancy data, and we demonstrate that even simple embedding approaches that utilize domain-specific metadata can improve the risk evaluation before and during the earlier pregnancy stages, comparatively to

probabilistic approaches. We also show that embedding approaches provide evidence about how feature values regarding each test individual are related to the risk labels, supporting explainability for its model prediction in such a way that humans understand.

Our embedding approach for risk assessment differs from previous works in three ways: (a) most of the open-domain embedding methods rely on previously seeing entities and relations during training in order to perform predictions in the test set – we consider our use case more realistic, when using a test set that corresponds to a new set of cases (only relations and object entities are seen in the training set, but not subjects); (b) we use a two-step approach that firstly embeds the test subjects into a pretrained embedding space, and then uses clustering to find entity similarities that can lead to a better prediction and explainability; finally, (c) although traditional embedding methods also rely on the labels being part of the embedding data, in our approach, the 'miscarriage' flag is never seen during training and validation, which we believe makes the embedding model reusable as input for multiple risk assessment criteria.

The ways we plan to extend this work include: (a) expanding current experiments to a wider pregnancy dataset (up to 20 years), reassessing risk year after year in order to draw the high-level picture on how clinical risk changes and how it is timely affected; (b) incorporating other embedding ontological constraints (e.g. disjoint sets, taxonomies and other hierarchical structure of classes, such as the ICD-10 organization), in addition to the typed-based entities, and evaluate how they can potentially improve the quality embedding representation of clinical data and subsequently improve risk assessment and further machine learning applications; and (c) validating how the proposed embedding approach would be effective when simultaneously applied to assess risk of multiple clinical conditions associated with high risk pregnancies, other than miscarriage, such as Hyperemesis gravidarum, gestational diabetes, and pre-eclampsia (toxemia).

References

1. Alves, N.C.d.C., Feitosa, K.M.A., Mendes, M.E.S., Caminha, M.d.F.C.: Complicações na gestação em mulheres com idade maior ou igual a 35 anos. Revista Gaúcha de Enfermagem **38** (2017)
2. Bearak, J., Popinchalk, A., Alkema, L., Sedgh, G.: Global, regional, and subregional trends in unintended pregnancy and its outcomes from 1990 to 2014: estimates from a bayesian hierarchical model. The Lancet. Global health **6**(4), e380–e389 (2018)
3. Bender Atik, R., Hepworth-Jones, B.E., Doyle, P.: Risk factors for miscarriage, p. 9–18. Cambridge University Press (2010)
4. Bona, C.: Avaliação de processos de software: Um estudo de caso em xp e iconix. Master's thesis, Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina (UFSC) (2002)
5. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K.Q. Weinberger (eds.) Advances in Neural Information Processing Systems 26, pp. 2787–2795. Curran Associates, Inc. (2013)
6. Borisov, N., Buzdin, A.: New paradigm of machine learning (ml) in personalized oncology: Data trimming for squeezing more biomarkers from clinical datasets. Frontiers in Oncology **9**, 658 (2019)
7. Branch, D.W., Heuser, C.: Recurrent Miscarriage, pp. 281–296. Springer New York, New York, NY (2010)
8. van den Broek, N.R.: Maternal and newborn health. In: E.T. Ryan, D.R. Hill, T. Solomon, N.E. Aronson, T.P. Endy (eds.) Hunter's Tropical Medicine and Emerging Infectious Diseases, 10th edn., pp. 133 – 140. Content Repository Only!, London (2020)
9. Chari, S., Qi, M., Agu, N.N., Seneviratne, O., McCusker, J.P., Bennett, K.P., Das, A.K., McGuinness, D.L.: Representing study populations in scientific literature in knowledge graphs. In: Proceedings of the Scientific Literature Knowledge Base (SLKB) Workshop at Automatic Knowledge Base Construction (AKBC) (2019)
10. Chung, M.W.H., Liu, J., Tissot, H.: Clinical knowledge graph embedding representation bridging the gap between electronic health records and prediction models. In: M.A. Wani, T.M. Khoshgoftaar, D. Wang, H. Wang, N. Seliya (eds.) 18th IEEE International Conference On Machine Learning And Applications, ICMLA 2019, Boca Raton, FL, USA, December 16-19, 2019, pp. 1448–1453. IEEE (2019). DOI 10.1109/ICMLA.2019.00237
11. Chung, M.W.H., Tissot, H.: Evaluating the effectiveness of margin parameter when learning knowledge embedding representation for domain-specific multi-relational categorized data. In: StarAI 2020 - Ninth International Workshop on Statistical Relational AI. AAAI (2020). URL http://www.starai.org/2020/
12. Collaborators, G..R.F.: Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. Lancet (London, England) **392**(10159), 1923–1994 (2018)
13. Curtis, C.: Meeting health care needs of women experiencing complications of miscarriage and unsafe abortion: Usaid's postabortion care program. Journal of Midwifery & Women's Health **52**(4), 368 – 375 (2007). Special Issue
14. Davis, J., Goadrich, M.: The relationship between precision-recall and roc curves. In: Proceedings of the 23rd International Conference on Machine Learning, ICML '06, p. 233–240. Association for Computing Machinery, New York, NY, USA (2006)
15. Diaz-Quijano, F.A., Pelissari, D.M., Chiavegatto Filho, A.D.P.: Zika-associated microcephaly epidemic and birth rate reduction in brazilian cities. American Journal of Public Health **108**(4), 514–516 (2018)
16. Drey, E.A., Foster, D.G., Jackson, R.A., Lee, S.J., Cardenas, L.H., Darney, P.D.: Risk factors associated with

presenting for abortion in the second trimester. Obstetrics & Gynecology **107**(1) (2006)

17. Durand, G., Bretelle, F., Fenollar, F.: Complications of pregnancy: Maternal perspectives. In: J. Cohen, W.G. Powderly, S.M. Opal (eds.) Infectious Diseases (Fourth Edition), 4 edn., pp. 498 – 504.e2. Elsevier (2017)

18. Ferri, J., Tissot, H., Del Fabro, M.D.: Integrating approximate string matching with phonetic string similarity. In: A. Benczúr, B. Thalheim, T. Horváth (eds.) Advances in Databases and Information Systems, pp. 173–181. Springer International Publishing, Cham (2018)

19. Gesese, G.A., Biswas, R., Sack, H.: A comprehensive survey of knowledge graph embeddings with literals: Techniques and applications. In: M. Alam, D. Buscaldi, M. Cochez, F. Osborne, D.R. Recupero, H. Sack (eds.) Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG2019) Co-located with the 16th Extended Semantic Web Conference 2019 (ESWC 2019), Portoroz, Slovenia, June 2, 2019, *CEUR Workshop Proceedings*, vol. 2377, pp. 31–40. CEUR-WS.org (2019)

20. Hogan, M.C., Foreman, K.J., Naghavi, M., Ahn, S.Y., Wang, M., Makela, S.M., Lopez, A.D., Lozano, R., Murray, C.J.: Maternal mortality for 181 countries, 1980–2008: a systematic analysis of progress towards millennium development goal 5. The Lancet **375**(9726), 1609–1623 (2010)

21. Johnson, J.M., Khoshgoftaar, T.M.: Survey on deep learning with class imbalance. Journal of Big Data **6**(1), 27 (2019)

22. Juric, D., Stoilos, G., Melo, A., Khodadadi, J.M.M.: A system for medical information extraction and verification from unstructured text. In: The Thirty-Second Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-2020). AAAI (2020)

23. de La Rochebrochard, E., Thonneau, P.: Paternal age and maternal age are risk factors for miscarriage; results of a multicentre European study. Human Reproduction **17**(6), 1649–1656 (2002)

24. Lam, J., Noor, Y.A., Supriyanto, E.: Ontology driven knowledge base for high risk pregnancy management. In: 2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), pp. 196–201 (2015)

25. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence, AAAI'15, pp. 2181–2187. AAAI Press (2015)

26. Lockwood, C.J.: Prediction of pregnancy loss. The Lancet **355**(9212), 1292–1293 (2000)

27. Medeiros, G.R., Nickel, D.A., Calvo, M.C.M.: Building a model to evaluate use of the National Program for Improving Primary Health Care Access and Quality in Brazil. Epidemiologia e Serviços de Saúde **28** (2019)

28. Mohamed, S.K., Nounu, A., Nováček, V.: Biological applications of knowledge graph embedding models. Briefings in Bioinformatics (2020)

29. Monteiro, C.A., Maia, E.G., Machado, I.E., Nico, L.S., Santos, M.A.S., de Fatima Marinho de Souza, M., de Freitas, M.G., Coelho, M.R.S., Campos, M.O., de Oliveira, P.P.V., Claro, R.M., Rodrigues, R., Bernal, R.T.I., de Araujo Andrade, S.S.C., da Silva, S.U.: Surveillance system for risk and protective factors for chronic diseases by telephone survey (Vigitel): changes in weighting methodology. Epidemiologia e Serviços de Saúde **26**, 701 – 712 (2017)

30. Nickel, M., Rosasco, L., Poggio, T.: Holographic embeddings of knowledge graphs. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16, pp. 1955–1961. AAAI Press (2016)

31. Nickel, M., Tresp, V., Kriegel, H.P.: A three-way model for collective learning on multi-relational data. In: Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML'11, pp. 809–816. Omnipress, USA (2011)

32. Njie, G.J., Proia, K.K., Thota, A.B., Finnie, R.K., Hopkins, D.P., Banks, S.M., Callahan, D.B., Pronk, N.P., Rask, K.J., Lackland, D.T., Kottke, T.E.: Clinical decision support systems and prevention. American Journal of Preventive Medicine **49**(5), 784–795 (2015). DOI 10.1016/j.amepre.2015.04.006. URL https://doi.org/10.1016/j.amepre.2015.04.006

33. Obermeyer, Z., Emanuel, E.J.: Predicting the future - big data, machine learning, and clinical medicine. The New England journal of medicine **375**(13), 1216–1219 (2016)

34. Rajkomar, A., Hardt, M., Howell, M.D., Corrado, G., Chin, M.H.: Ensuring fairness in machine learning to advance health equity. Annals of internal medicine **169**(12), 866–872 (2018)

35. Saravelos, S.H., Li, T.C.: Unexplained recurrent miscarriage: how can we explain it? Human Reproduction **27**(7), 1882–1886 (2012)

36. Small, M.J., James, A.H., Kershaw, T., Thames, B., Gunatilake, R., Brown, H.: Near-miss maternal mortality: Cardiac dysfunction as the principal cause of obstetric intensive care unit admissions. Obstetric Anesthesia Digest **33**(1) (2013)

37. Sutton, R.T., Pincock, D., Baumgart, D.C., Sadowski, D.C., Fedorak, R.N., Kroeker, K.I.: An overview of clinical decision support systems: benefits, risks, and strategies for success. npj Digital Medicine **3**(1) (2020). DOI 10.1038/s41746-020-0221-y. URL https://doi.org/10.1038/s41746-020-0221-y

38. Tcheng, J.E., Bakken, S., Bates, D.W., Bonner III, H., Gandhi, T.K., Josephs, M., Kawamoto, K., Lomotan, E.A., Mackay, E., Middleton, B., Teich, J.M., Weingarten, S., Lopez, M.H.: Optimizing Strategies for Clinical Decision Support - Summary of a Meeting Series. National Academy of Medicine, Washington, DC (2017)

39. Tissot, H.: HEXTRATO: Using Ontology-based Constraints to Improve Accuracy on Learning Domain-specific Entity and Relationship Embedding Representation for Knowledge Resolution. In: IC3K 2018 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, vol. 1, pp. 72–81. SciTePress (2018)

40. Tissot, H., Dobson, R.: Combining string and phonetic similarity matching to identify misspelt names of drugs in medical records written in portuguese. Journal of Biomedical Semantics **10**(1), 17 (2019)

41. Trouillon, T., Welbl, J., Riedel, S., Gaussier, E., Bouchard, G.: Complex embeddings for simple link prediction. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning, *ICML*, vol. 48, p. 2071–2080. JMLR.org (2016)

42. Wang, Q., Mao, Z., Wang, B., Guo, L.: Knowledge graph embedding: A survey of approaches and applications. IEEE Transactions on Knowledge and Data Engineering **29**(12), 2724–2743 (2017)

43. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: C.E. Brodley, P. Stone (eds.) Proceedings of the Twenty-Eighth

AAAI Conference on Artificial Intelligence, pp. 1112–1119. AAAI Press (2014)

44. Wasylewicz, A.T.M., Scheepers-Hoeks, A.M.J.W.: Clinical decision support systems. In: Fundamentals of Clinical Data Science, pp. 153–169. Springer International Publishing (2018). DOI 10.1007/978-3-319-99713-1_11. URL https://doi.org/10.1007/978-3-319-99713-1_11

45. Xu, Z., Zhang, J., Zhang, Q., Yip, P.S.F.: Explainable learning for disease risk prediction based on comorbidity networks. In: 2019 IEEE International Conference on Systems, Man and Cybernetics, SMC 2019, Bari, Italy, October 6-9, 2019, pp. 814–818 (2019). DOI 10.1109/SMC.2019.8914644

46. Yang, Y., Fang, Y., Wang, X., Zhang, Y., Liu, X., Yin, Z.: A retrospective cohort study of risk factors and pregnancy outcomes in 14,014 chinese pregnant women. Medicine **97**(33), e11748 (2018)

47. Zikos, D., DeLellis, N.: CDSS-RM: a clinical decision support system reference model. BMC Medical Research Methodology **18**(1) (2018). DOI 10.1186/s12874-018-0587-6. URL https://doi.org/10.1186/s12874-018-0587-6