Mother: a maternal online technology for health care dataset

Odongo Steven Eyobu[1,2]*, Brian Angoda Nyanga[1,2], Lukman Bukenya[1,2], Daniel Ongom[1,2] and Tonny J. Oyana[1]

Abstract

Objectives These data enable the development of both textual and speech based conversational machine learning models that can be used by expectant mothers to provide answers to challenges they face during the different trimesters of their pregnancy. Such models are key to the improvement of the lives of pregnant mothers, specifically in low resourced settings where doctors advise is limited by access to hospitals and language barrier. These data were used to develop a conversational chatbot model tailored for mothers in their first, second and third trimesters of pregnancy.

Data description 503 question and answer pairs on maternal health were collected through a survey of challenges facing pregnant mothers in a rural and semi-urban area of Uganda. The answers to the questions were provided and validated by professional medical personnel. The participants were purposively sampled, focusing on women in their 1st, 2nd and 3rd trimesters, with a 94% response rate. The dataset addresses common health concerns, symptoms, and conditions associated with pregnancy, particularly for women without immediate access to medical personnel. It targets maternal health outcomes such as pregnancy, morbidity, and mortality, specifically among women of reproductive age.

Keywords Maternal health, Pregnancy, Question and answer knowledge base, Electronic health (E-health), Conversational chatbots

Objective

According to a 2019 UNICEF report [1], women in sub-Saharan Africa are fifty times more likely to die from childbirth than women in high-income countries. This is attributed to various factors including limited access to healthcare facilities and professional doctors, lack of emergency care, healthcare information and malnutrition. This article presents a comprehensive question-and-answer dataset [2] for maternal health, designed to enable the development of a conversation chatbot that provides healthcare information to support pregnant mothers with limited access to doctors whenever needed particularly in resource constrained rural areas. The knowledge base [2] provides answers to various challenges of maternal health, including prenatal care, nutrition, pregnancy complications, childbirth, postpartum care, and maternal wellness. With approximately 800 women dying daily due to pregnancy-related complications in rural Africa [3], this dataset [2] serves as a foundation for building intelligent data-driven conversational models to run a chatbot that directly supports pregnant mothers with instant, immediate and accurate

*Correspondence:
Odongo Steven Eyobu
odongo.eyobu@mak.ac.ug
[1]Geospatial Data and Computational Intelligence lab, College of Computing and Information Science, Makerere University, P.O Box, 7062, Kampala, Uganda
[2]Department of Networks, School of Computing and Information Technology, Makerere University, P.O Box, 7062, Kampala, Uganda

Eyobu *et al. BMC Research Notes*   (2025) 18:150

Page 2 of 3

Table 1 Data file descriptions

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)	Reference
Data file 1	intents	Excel (.csv)	Harvard Dataverse https://doi.org/10.7910/DVN/EZLCH3	[2]
Data file 2	patterns	Excel (.csv)	Harvard Dataverse https://doi.org/10.7910/DVN/EZLCH3	[2]
Data file 3	responses	Excel (.csv)	Harvard Dataverse https://doi.org/10.7910/DVN/EZLCH3	[2]
Data file 4	mother_question_and_answer_pairs_data	JSON	Harvard Dataverse https://doi.org/10.7910/DVN/EZLCH3	[2]
Data file 5	mother_intents_patterns_responses_data	JSON	Harvard Dataverse https://doi.org/10.7910/DVN/EZLCH3	[2]

on-demand information concerning their healthcare needs. It is envisaged that the resultant conversational chatbots will be able to receive queries and deliver health care information in preferred local languages to take care of the language diversities of rural populations that do not understand English particularly in sub-Saharan Africa. By empowering expectant mothers with reliable information, particularly in low- and middle-income countries, there is a great opportunity to contribute to a reduction in maternal mortality and improving maternal health outcomes.

Data description

Questions on challenges and lifestyle during pregnancy from 500 expectant mothers were compiled and answers to each question were provided by medical professionals to formulate the question-and-answer pair textual dataset. The age range of the expectant mothers was 20–50 years. The data collected clearly shows pregnancy challenges associated with women in rural settings. These are mainly associated with nutrition challenges, antenatal care, and postpartum care.

Participants were purposively sampled, focusing on women in their 1st, 2nd and 3rd trimesters, with a 94% response rate. 161 mothers were in their first trimester, 142 mothers were in their second trimester, 197 mothers were in their 3rd trimester.

After collecting the data, preprocessing steps were performed to realign the questions and answers into clear English sentences to enhance readability of the texts. Question and answer pairs that were similar were grouped to form patterns and responses respectively. The patterns represented the questions that have the same meaning. The responses represented the answers that had the same meaning. These patterns and responses were further given tags and contexts. The tags were used to show the general topic that the questions and answers addressed.

The contexts on the other hand represented the specific topic that the questions and answers were addressing. Intents were then created by grouping tags, contexts,

patterns and responses. This allowed normal question and answer pairs to be modeled into a dataset that can be used to train a BERT [4] model for a chat bot. The reason for having multiple patterns for some questions is because a user interacting with a chat bot may ask a question in different ways and hence the model would have to decipher exactly what the user wants to know and be able to give an appropriate response. Also, in cases where a response could be framed in different ways, multiple responses were created to provide a variety of answers to a given question. Questions that contained single-word answers were rephrased by adding more words to elaborate the answer so that a user could better understand the output from the model.

Data file 1 [2] is: Intents (1) id, (2) tag, (3) context_set where id is the unique identifier for an intent, tag is a general topic about the intent and context_set is a specific topic about the intent.

Data file 2 [2]: Patterns (1) id, (2) intent_id, (3) content, where id is the unique identifier of a pattern, intent_id represents the intent that the pattern belongs to and content is the actual question.

Data file 3 [2]: Responses (1) id, (2) intent_id, (3) content, where id is the unique identifier of a response, intent_id represents the intent that the response belongs to and content is the actual answer.

Data file 4 [2]: Question and answer pairs.

Data file 5 [2]: Question and answer pairs transformed into intents, patterns and responses (Table 1).

Limitations

The dataset [2] is presented in English and therefore requires a translator in order for it to be used in English language constrained populations of expectant mothers.

The dataset [2] is limited to responses from only 500 participants, which may not fully capture the diverse challenges faced by pregnant mothers across various regions and demographics. So, the repository will undergo continuous updates bimonthly.

Applying or use of the dataset for conversational chatbots is not a total replacement of doctors but works as

an emergency information dissemination tool in resource constrained areas.

Declarations

Ethics approval and consent to participate
This research involved human subjects specifically to provide questions on issues affecting their maternal health. This research work was approved by the Mildmay Uganda Research & Ethics Committee (MUREC) protocol number **MUREC-2023-243** and the Uganda National Council of Science and Technology. Informed consent to participate was obtained from all of the participants in the study.

Competing interests
The authors declare no competing interests.

References
1. Rafanelli A. accessed Dec. 11, The Risks to Pregnant Women in Sub-Saharan Africa: They're Focused on Just Getting Through It., *Direct Relief*. https://www.directrelief.org/2021/12/the-risks-to-pregnant-women-in-sub-saharan-africa-n-theyre-focused-on-just-getting-through-it/ (2024).
2. Eyobu OS, Daniel O, Angoda B, Bukenya, Lukman TJ, Oyana. 2024, MOTHER: A dataset for maternal online technology for Health Care Dataset, https://doi.org/10.7910/DVN/EZLCH3, Harvard Dataverse, V4.
3. World Health Organization: WHO. Maternal mortality, Apr. 26, 2024. https://www.who.int/news-room/fact-sheets/detail/maternal-mortality
4. Devlin J. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.