

Proper Dataset Documentation

“Early Pregnancy Detection and prevention using Advanced machine learning Algorithms”

1.Dataset Name and Description:

- Dataset Name: Association of Vitamin D Deficiency with Early Pregnancy Loss
- Description: The dataset focuses on assessing the correlation between vitamin D deficiency and early pregnancy loss, analyzing demographic, obstetric, anthropometric, and biochemical variables among participants.

2. Dataset Creation:

- Origin: The dataset was created through a case-control study conducted at Rangpur Medical College & Hospital, Bangladesh.
- Methodology: Data was collected through purposive sampling of patients attending the Outpatient and Indoor Department of Obstetrics and Gynaecology.
- Time Period: The study was carried out from September 2020 to August 2021.
- Team and Contributors: The study involved a team at the Institute of Child and Mother Health which is under the rangpur medical college & hospital. Although specific names and roles are not mentioned in the document [70†source] .

3. Data Characteristics:

- Variables: The dataset includes demographic, obstetric, anthropometric, and biochemical variables, with early pregnancy loss as the dependent variable and factors like serum vitamin D level, age, BMI, and sun exposure as independent variables [73†source] .

- Units: Serum vitamin D levels are measured in ng/ml.
- Data Types: The dataset includes categorical and numerical data types.
- Sample Size: 34 women with miscarriage in the first trimester and 34 women with normal pregnancy in the first trimester 【72†source】 .

4. Data Collection Process:

- Selection Criteria: The study included women with miscarriages and normal pregnancies during the first trimester, with specific inclusion and exclusion criteria 【72†source】 .
- Data Collection Instruments: A structured questionnaire was used to collect relevant data 【74†source】 .
- Data Collection Procedures: Participants provided detailed personal, family, and medical history, and underwent clinical examination and blood tests for vitamin D estimation.
- Quality Control: At this stage there are different statistical test has been done. The how data quality has been ensured as data validation.

5. Data Confidentiality and Ethic:

- Consent: Informed written consent was obtained from all participants.
- Anonymization: Research data was coded, and privacy was maintained.
- Ethical Considerations: Ethical clearance was obtained from the Institutional Review Board, and the study maintained ethical standards throughout 【75†source】 .

6. Usage and Access:

- License: Not specified in the document.
- Access: Not specified in the document.
- Citation: Not specified in the document.

7. Acknowledgments:

- Funding Sources:
The authors gratefully acknowledged the support of the Ministry of Science and Technology of China (grant number G2021171024L), the High-Level Talent Project of Chang'an University (grant number 300203110029), the innovation creative base

project of Shaanxi Province (grant number 300201000173), the Bangladesh Bureau of Educational Information & Statistics (BANBEIS), the Ministry of Education, the Government of the People's Republic of Bangladesh (PCN No. IC20201325), and the participants in the test.

- Collaborating Institutions: The study was conducted at the Institute of Child and Mother Health, Dhaka.

8. Change and Version Control:

- Version: Not specified in the document.
- Change Log: Not specified in the document.

9. Data Presentation:

- Data were presented in appropriate tables 【76†source】 .
- A study design flow chart was included 【77†source】 .

This documentation summarizes key aspects of the dataset based on the information provided in the thesis.