

Data Collection and Preprocessing Phase

Date	23 June 2025
Name	Aniket Jingonda Patil
Project Title	Health Classification System
Maximum Marks	2 Marks

Data Collection Plan

Section	Description
Project Overview	Develop a health classification system to assist healthcare providers and expectant parents in assessing fetal health using cardiotocography (CTG) data. By analyzing 21 CTG features, this project aims to provide accurate predictions for Normal, Suspect, or Pathological categories, enhancing prenatal care.
Data Collection Plan	The dataset used for this project was sourced from the UCI Machine Learning Repository and contains 2126 records with 21 features (e.g., baseline value, accelerations, histogram).
Raw Data Sources Identified	The raw data for this project was obtained from the UCI Machine Learning Repository dataset titled “Cardiotocography Data Set” by Ayres de Campos et al. The dataset is publicly available at https://archive.ics.uci.edu/ml/datasets/Cardio to cography and includes key CTG-related attributes such as baseline value, accelerations, fetal movements, uterine contractions, and fetal health labels (1=Normal, 2=Suspect, 3=Pathological).

Raw Data Sources

Source Name	Description	Location/URL	Format	Size	Access Permissions
SmartInternz Provided Dataset	<p>This dataset contains 2126 records of features extracted from Cardiotocogram exams, which were then classified by three expert obstetricians into 3 classes:</p> <ul style="list-style-type: none"> • Normal • Suspect • Pathological 	Fetal AI dataset	CSV	~ 46KB	Public