Data Collection and Preprocessing Phase

Date	27th July 2024
Team ID	740076
Project Title	FETAL AI:USING MACHINE LEARNING TO PREDICT AND MONITOR FETAL HEALTH
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

The data collection plan for the Fetal AI system focuses on obtaining high-quality ultrasound images and comprehensive clinical records from partnering hospitals and medical imaging centers. Raw data sources include direct feeds from ultrasound machines and electronic health records, ensuring a continuous and diverse influx of data for robust model training. Key partnerships and standardized data-sharing agreements will be established to facilitate reliable data acquisition and ensure consistency across all data sources.

Data Collection Plan Template

Section	Description
Project Overview	The Fetal AI project aims to enhance fetal health assessment using machine learning models that analyze ultrasound images and clinical data.
Data Collection Plan	Collect high-resolution ultrasound images and detailed clinical records from partner hospitals and imaging centers for comprehensive data analysis.
Raw Data Sources	Data is sourced from direct feeds of ultrasound machines and electronic health record systems to ensure a robust dataset for training and evaluation.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Hospital A	High-resolution ultrasound images and associated clinical records from a regional medical center.	Secure internal database	DICOM, CSV	500 GB	Restricted access, NDA required
Imaging Center B	Ultrasound imaging data from prenatal screenings, including patient demographics and outcomes.	https://imagingcenterb.com	JPEG, JSON	250 GB	Requires user authentication

Health Network C	Aggregated electronic health records from multiple clinics, focusing on prenatal care and outcomes.	Internal health network cloud	HL7, XML	1 TB	Permission from health network admin
Open Data Portal	Publicly available fetal health datasets for training and benchmarking AI models.	https://data.fetalhealth.org	CSV	100 GB	Open access