Data Collection and Preprocessing Phase

Date	12 June 2025
Name	Sakshi Santosh Patil
Project Title	Fetal Health Classification System
Maximum Marks	6 Marks

Data Preprocessing

The data will be preprocessed by scaling, handling class imbalance, detecting outliers, and adjusting target labels. These steps will enhance data quality, pro mote model generalization, and improve convergence during XGBoost training, ensuring robust performance for fetal health classification.

Section	Description
Data Overview	The dataset (fetal_health.csv) contains 2126 records with 21 CTG features (e.g., baseline value, accelerations, histogram
Scaling	All 21 CTG features are scaled using StandardScaler to standardize values (mean=0, variance=1), saved as scaler.pkl.
Handling Class Imbalance	Stratified sampling is used during train-test split (80%-20%) to preserve class distribution. Class weights are applied in XGBoost to address imbalance.
Outlier Detection	Outliers in features (e.g., baseline value) are identified using the IQR method and capped or removed after validation..
Target Adjust ment	The target variable (fetal_health) is adjusted from 1,2,3 to 0,1,2 for compatibility with XGBoost
Edge Detection	Not applicable for text data.
Color Space Conversion	Not applicable for text data.

Image Cropping	Not applicable for text data.
Batch Normalization	Not applicable for text data.

Data Preprocessing Code Screenshots

| Loading Data | ```python
from flask import Flask, request, jsonify, render_template
import joblib
import numpy as np

app = Flask(__name__)

# Load the model and scaler
model = joblib.load('fetalai_model.pkl')
scaler = joblib.load('scaler.pkl')
``` |
| Resizing | *Not applicable* |
| Data Mapping | ```python
# Map prediction to a label
labels = {1: "Normal", 2: "Suspect", 3: "Pathological"}
result = labels[prediction]
message = "Normal fetal health." if prediction == 1 else "Potential concern detected. Consult a healthcare provider."
``` |
Data Augmentation	*Not applicable*
Denoising	*Not applicable*
Edge Detection	*Not applicable*
Color Space Conversion	*Not applicable*

Image Cropping	*Not applicable*
Batch Normalization	*Not applicable*