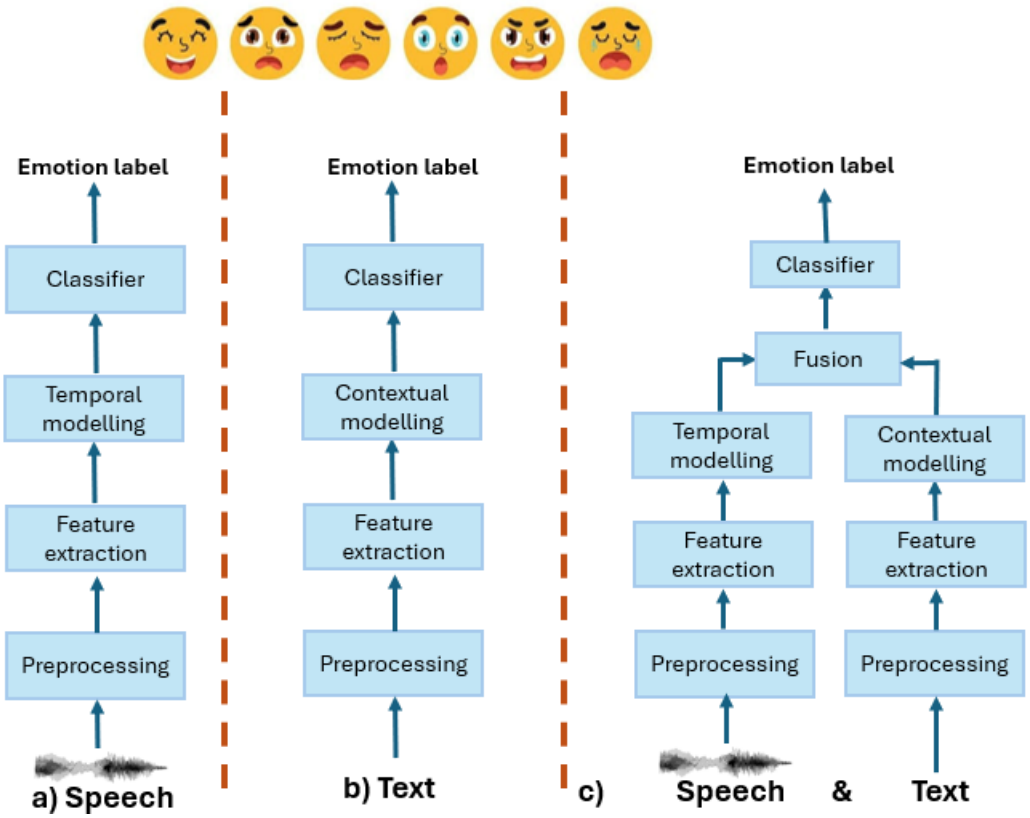
Assignment 2

Multimodal Emotion Recognition

Objective: Build a system that recognizes emotions using only speech, only text and combination of both (Multimodal) as inputs. The functional blocks to build the system are provided in the figure below, you are required to choose appropriate ML/DL architectures for each block.



Dataset: Toronto emotional speech set (TESS) - link (Available on Kaggle)

- It contains speech samples along with corresponding transcripts/text and emotion labels.
- Please go through the dataset description provided in the link.

Role of Functional blocks:

1. Preprocessing

- **Speech:** Handle varying lengths/quality, choose sampling rate, trim silence
- **Text:** Cleaned, tokenized text

2. Feature Extraction

- **Speech:** Extract emotional cues (time_steps × features)
- **Text:** Extract emotional cues (tokens × features)

3. Temporal/contextual Modelling

- **Speech:** Learn emotional pattern over time.
- **Text:** Learn emotional meaning across tokens in context.

4. Fusion

- Combine representations from both modalities (Speech & Text) to obtain a unified representation for emotion classification.

5. Classifier

- Predict the emotion label from the learned representations produced by the previous block.

Deliverables

1. Code

```
project/
├── models/
│   ├── speech_pipeline/
│   │   |── train.py
│   │   |── test.py
├   ├── text_pipeline/
│   │   |── train.py
│   │   |── test.py
│   └── fusion_pipeline/
|       |── train.py
|       |── test.py
|
├── Results/
|   |── All 3 model variants accuracy tables
```

```
|   └── plots
├── README.md
└── requirements.txt
```

2. Report

A. Architecture Decisions - For each block: What architecture? And why?

B. Experiments - Speech-only, Text-only, Multimodal: Comparison

C. Analysis

- Which emotions are easiest/hardest to classify? Why?
- When does fusion help most?
- Error analysis: 3-5 failure cases
- Visualize the separability of emotion clusters using the learned representations from:
    - Temporal Modelling block
    - Contextual Modelling block
    - The Fusion block

3. GitHub Repository: Push all scripts, setup instructions, report and results to a public repository. Ensure all links (GitHub and Drive) are publicly accessible for evaluation.