



# MIDAS@IIITD

Multimodal Digital Media Analysis Lab

## MIDASS@IIITD Internship Task

### GENERAL INSTRUCTIONS

Failing to follow any of the instructions below will lead to rejection of your submission.

1. This task has 3 problems
  - a. Speech Emotion Problem (This problem is mandatory, details given below).
  - b. [An image segmentation problem.](#)
  - c. [A tabular problem.](#)
2. [Google Form Submission Link](#)
3. **The audio problem is mandatory, in addition pick any of the two other problems to solve.** You may solve all 3, however, only two will be evaluated.
4. For the Kaggle challenges - submit your solution to their respective kaggle leaderboards and give us your kaggle username in the google form [here](#).
5. In addition, submit the link to a github repository with jupyter notebooks or scripts for the problem solution. The repository should be organized with sub-folders, one folder for each problem.
6. Do not commit any data files for the problem to the repository.
7. You need to include a bash script/dockerfile that creates an appropriate conda environment/docker container to be able to run the code.
8. You have **15 days (Midnight, 30th October Anywhere on Earth)** to submit the solutions. No extensions will be provided.
9. Make sure your codes is properly documented. We recommend the following:-
  - a. Before each code block have a markdown block/docstring which mentions the following
    - i. What the code block is doing.
    - ii. What is your intuition behind doing this? Why do you think it is useful?
  - b. Keep an experiment log - document everything that worked or failed. This document(preferably a jupyter notebook) should be a snapshot of the process you follow to solve each problem.

10. You are not allowed to add any new data samples from an online source or self-labeled to the provided dataset (Just to make it clear, data augmentation techniques while learning is allowed, adding completely new samples from any external sources is not allowed).
11. If you have any doubts feel free to email at [midas@iiitd.ac.in](mailto:midas@iiitd.ac.in) or [hitkuli@iiitd.ac.in](mailto:hitkuli@iiitd.ac.in)

## Speech Emotion Problem

**Problem Statement** - Classical Classification problem

**Paper** - <https://github.com/SenticNet/MELD.git>

**Dataset** -

<https://drive.google.com/drive/folders/1fs0I7x8-BV-ZR13M19qG0YWcuZ5k7oZD?usp=s>  
[haring](#)

Dataset contains two folders train and valid. Individual folder contains **speech utterances/.wav files** of 5 categories on which to do classification.

**Good to have:** Visualizations, justifications, and reasoning behind your thought process, and your solution.

### Submission

Add a folder in the repository with the name `SPEECH EMOTION PROBLEM`.

This folder must contain anything necessary to run the code(except data) and a **bash script which creates a conda environment and download all the necessary packages required** and weights if using a DL/ML technique.

Must contain a **README.txt** file if not then **submission is nullified**. The folder must contain a **jupyter notebook**.

Apart from jupyter notebook the folder must contain a **.py function which takes input**, the path to the **test data folder** and return a **text file** of the accuracy for the same.

**Test data folder** structure: It contains .wav files only of all the classes in the one folder only.

**Text file** Structure: should contain the file name and the prediction with comma separated and each test case separated with “\n” (**next line**).

Sample **text file** can be:

[https://docs.google.com/document/d/1pJ8U8IZlGBGHHv\\_fZq6lqo9uwRGQTNRUGpt8-r6mcxU/edit?usp=sharing](https://docs.google.com/document/d/1pJ8U8IZlGBGHHv_fZq6lqo9uwRGQTNRUGpt8-r6mcxU/edit?usp=sharing)