MemoTag submission

In this task I have used the dataset from **Dementia bank**
It mainly contains 3 types of data in form of the audio HC , MCI , Dementia
HC – highly cognitive

MCI – Midly cognitive impairment

Dementia – Confirm dementia

although it also contains pre provided transcriptions
I decided to make my own audio transcriptions using whisper
I have already tried using pre provided transcriptions and gives better results but since the question asked to use majorly acoustic features rather than linguistic features from transcriptions so I made my own transcriptions

In this for the acoustic features I have used **saliero vad** to give me the number of pauses taken during the audio of the person
IMPROVEMENT – One could make a transcription giving the **time stamp of each and every word** just after it in the transcription itself or make a csv for the same for each audio and then use this data to determine how long a pause and how repetitive they are …

In this for making the transcriptions we could use the **higher versions of the whisper** which are computationally expensive but tend to give more accurate results for each word

There are various methods to extract acoustic features that could be used such as MFCC or some other ways such as **opensmile feature extraction toolkit** or **ege maps set of features** which can be used to get more **acoustic features**

One can later use these **acoustic and linguistic features together** and then use SOTA LLM such as **RoBERTa** rather than some simpler models which are known for their NLP tasks …

Saket Nandedkar

IIT Kanpur
Mechanical Engineering