**Clustering-Based Speech Emotion Recognition by Incorporating Learned Features and Dee**

**Introduction:**

Automatic recognition and identifification of emotions from speech signals in speech emotion recognition (SER) using machine learning is a challenging task [1]. SER is a quick and usual method of communication and exchanging information among humans and computers and has many real world applications in the domain of Human-computer interaction (HCI). Currently, researchers are facing a major challenge in feature extraction i.e., how to select a robust method to extract salient and discriminative features from speech signals to represent the emotional state of a speaker from

their acoustic contents.

Data Description:

The IEMOCAP [49] is a well-known dataset which is commonly used for recognition of emotional speeches, which has two types of dialogs, scripted and improvised. The dataset

consists of 10 experienced actors to records 12 hours of audiovisual data including audio, videos, motion of faces, speech and text transcriptions. The IEMOCAP dataset has

fifive sessions and each session consists of 2 actors (one male and one female) to record the emotional script with 3 to 15 second long with a 16 kHz sampling rate. Each session

has different categories of emotions like; anger, sad, happy, neutral, surprise, disgust, frustrated, excited and fearful which is annotated by three expert persons. Individually labeled the data, we select those utterances that two experts are agreed upon them.

Data Collected from the github multiple locations

**Existing Methods:**

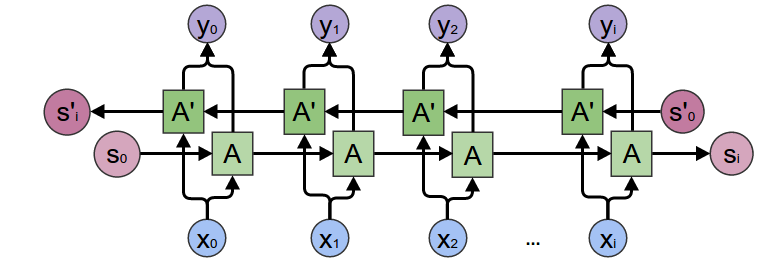
Bi LSTM model

**Proposed Method:BI LSTM with increased accruacy**

**Bi-LSTM:(Bi-directional long short term memory):**Bidirectional recurrent neural networks(RNN) are really just putting two independent RNNs together. This structure allows the networks to have both backward and forward information about the sequence at every time step

Using bidirectional will run your inputs in two ways, one from past to future and one from future to past and what differs this approach from unidirectional is that in the LSTM that runs backward you preserve information from the future and using the two hidden states combined you are able in any point in time to preserve information from both past and future.

IMG_256



**Steps for Deep Learning Algorithms:**

1. Install Anaconda Latest Version
2. Open anaconda Prompt
3. Conda create -n tf python=3.7
4. Conda activate tf
5. Install require softwares

**t**ensorflow==1.14.0

ipykernel==5.3.4

scikit-image==0.17.2

scikit-learn==0.23.2

pandas==1.1.1

matplotlib==3.3.1

Keras==2.3.1

Pillow==7.2.0

plotly==4.10.0

opencv-python==4.4.0.42

spacy==2.3.2

lightgbm==3.0.0

mahotas==1.4.11

matplotlib==3.3.1lightgbm==3.0.0

mahotas==1.4.11

nltk==3.5

matplotlib==3.3.1

xgboost==1.2.0

Jupyter

1. Activate environment for jupyter notebook(For execute the in jupter notebook)

python -m ipykernel install --user --name=

1. Goto project Directory

Note: For Text related project. Need to Download

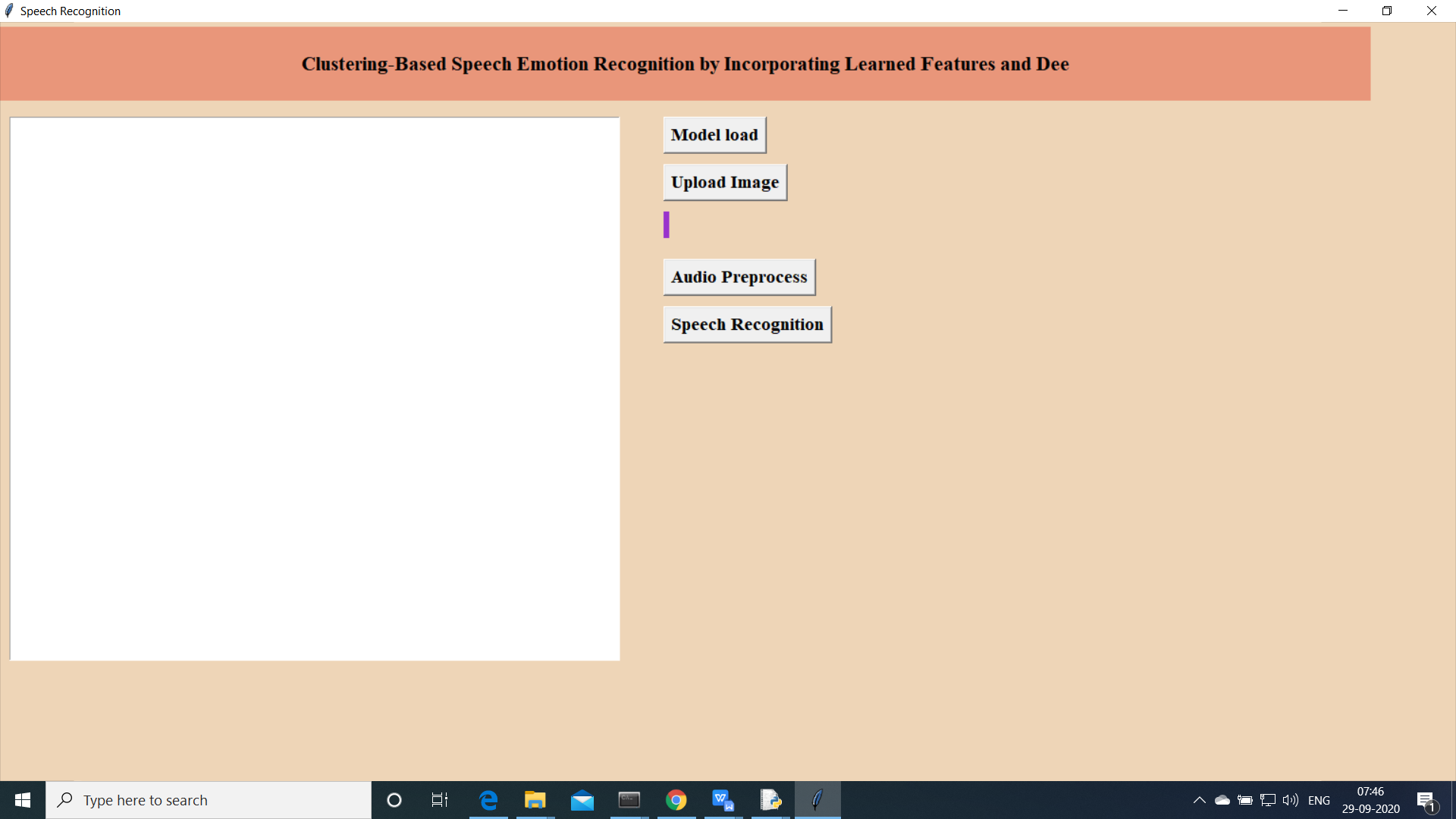
1. Open anaconda Prompt
2. Python
3. Import nltk
4. Nltk.download()

**Project Development Modules:**

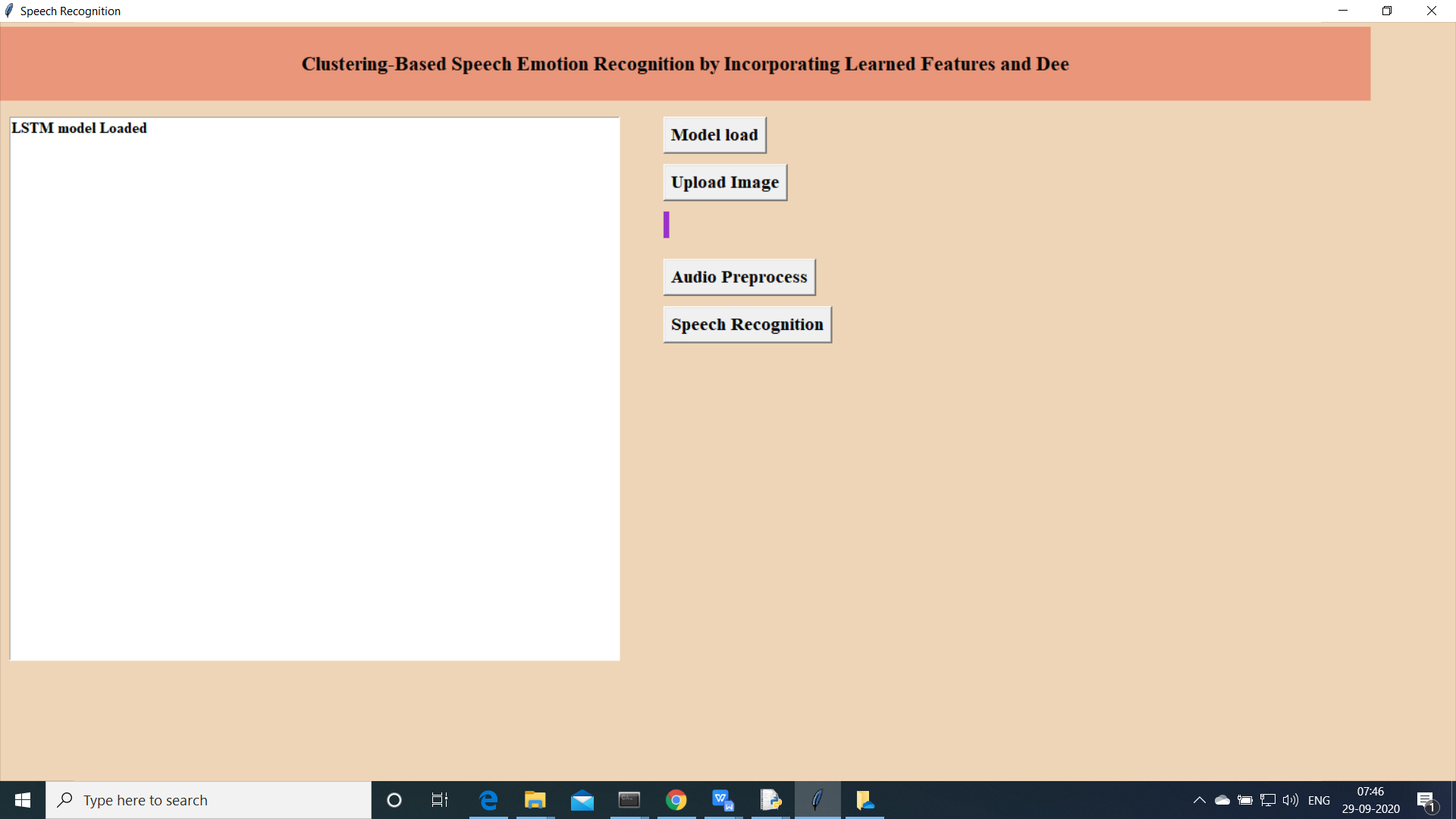
1. **Data Collection:** Collect sufficient data samples and legitimate software samples. 
2. **Audio feature extraction** :Audio features extraced from the speech of different actors
3. **Train and Test Modelling: Split the data into train and test data Train will be used for training the model and Test data to check the performace**
4. **Modelling:** BILSTM model build and model is saved
5. **Predict Select an single image and do basic image processing and predict using BILSTM model**

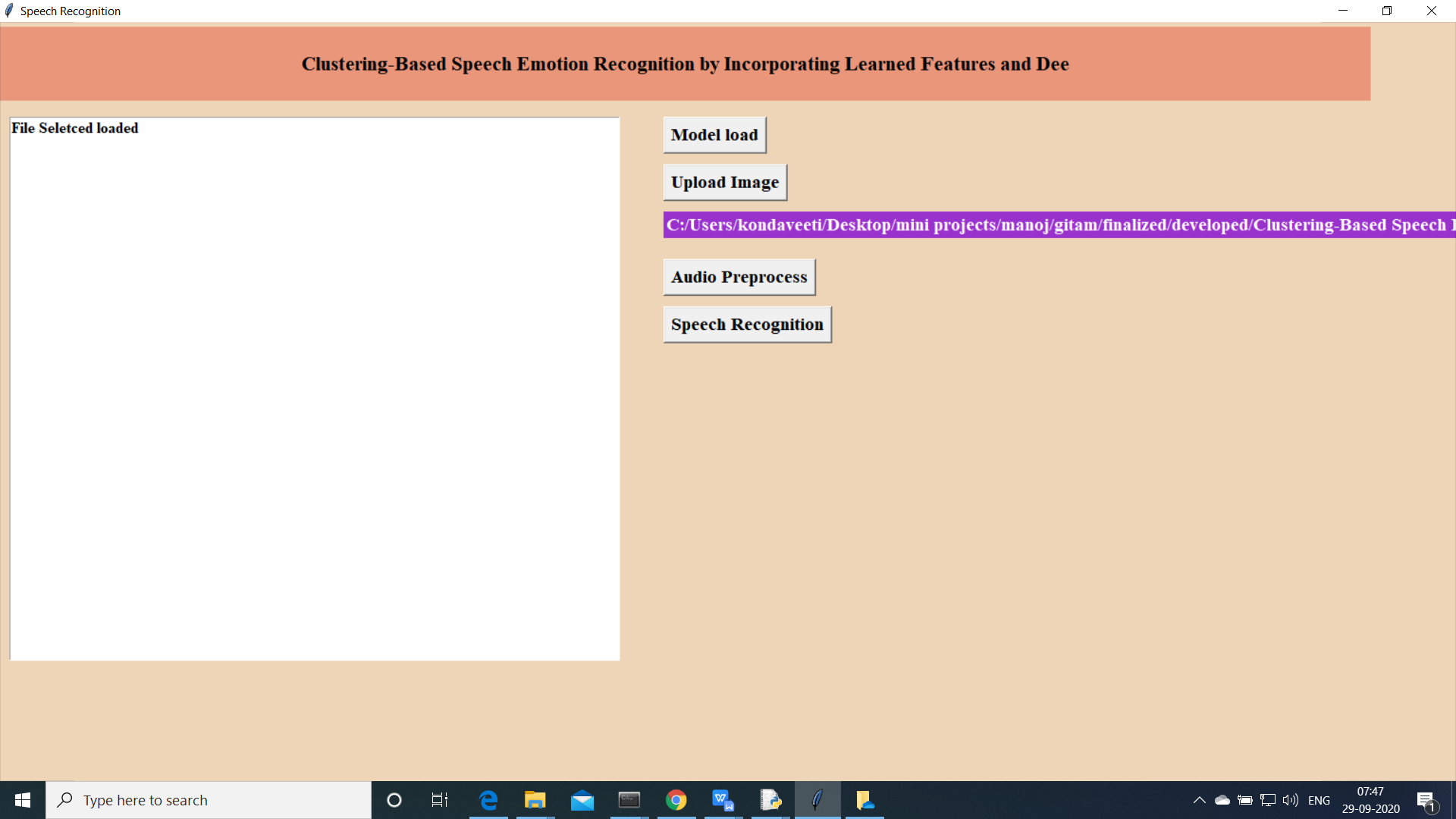
**Execution Steps:**

1. Open anaconda Prompt
2. Conda activate tf
3. Goto Project Directory
4. Python final.py

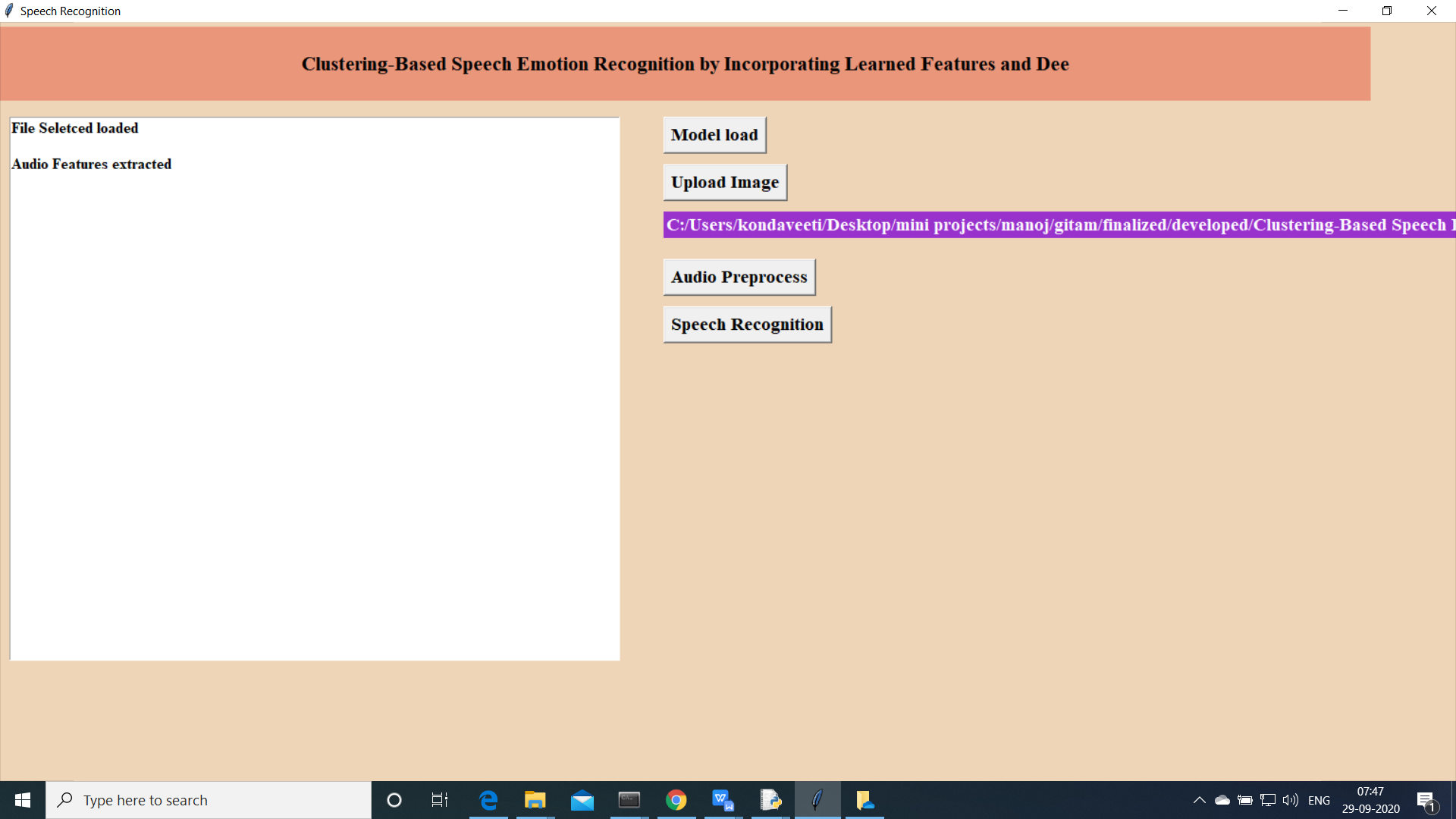


Above screen will be opened.

1. Load the lodel
2. 
3. Select an Audio to be predicted using trained model



Audio Features extracted



1. Predict speech recognition using model

