A Novel Speech Intelligibility Enhancement Model based on Canonical Correlation and Deep Learning

Audio files often comes with noises and there are many algorithms exists to remove noise from audio files. All existing algorithms were utilizing distance between clean and noise audio to replace noise audio with clean audio by utilizing close distances. Often this existing algorithms give imperfect or undesirable results. To overcome from above issue author of this paper employing Canonical Correlation Deep Learning based algorithm to enhance or improve speech quality.

Canonical correlation based short-time objective intelligibility (CC-STOI) cost function is utilized to train a fully convolutional neural network (FCN) model. Canonical correlation analysis (CCA) is a statistical method that measures the relationship between two sets of variables. It's a multivariate analysis of correlation that analyzes latent variables, which are not directly observed, that represent multiple variables, which are directly observed. CCA analyse clean and noisy pairs of dataset and then replace noisy part with clean features to remove noise from audio.

To train propose CC-STOI algorithm author has used CLEAN and NOISE dataset bot not publish on internet so we utilize and Clean and Noise audio dataset from Microsoft. Fully Connected CC-STOI propose algorithm get trained on NOISE and CLEAN audio features and when we input NOISY audio then propose algorithm will replace noisy features with clean features to enhance audio.

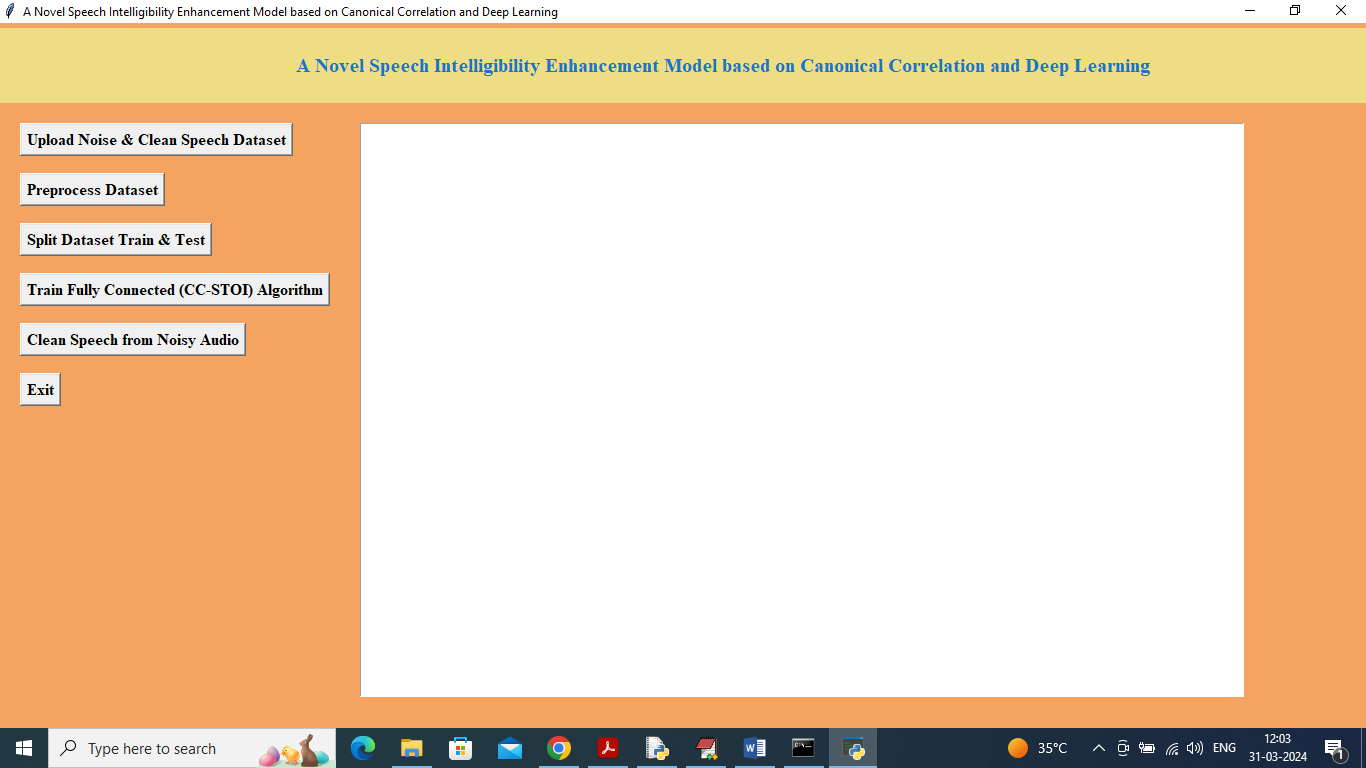
Propose algorithm performance is evaluated in terms of Mean Square Error (MSE) which refers to difference between TRUE and predicted features so the lower the MSE the better is the algorithm.

To implement this project we have designed following modules

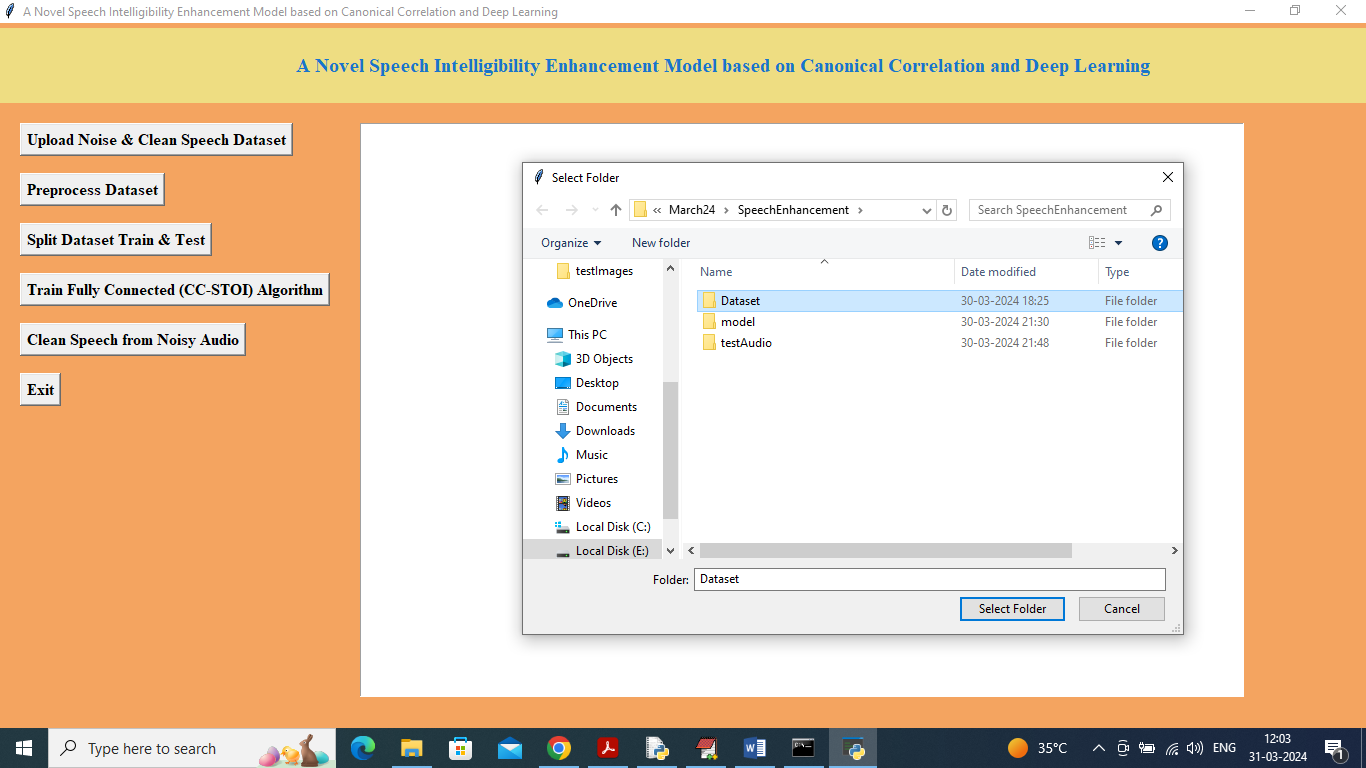
1. Upload Noise & Clean Speech Dataset: using this module will upload audio dataset to application and then application will extract features from CLEAN and NOISE audio and then generate a training array
2. Pre-process Dataset: all extracted audio features will be reshape as per Fully connected deep learning algorithm and then normalize features values
3. Split Dataset Train & Test: processed audio features will be split into train and test where application will be using 80% dataset for training and 20% for testing
4. Train Fully Connected (CC-STOI) Algorithm: 80% training features will be input to propose algorithm to train a model and this model will be applied on 20% test data to calculate prediction MSE error
5. Clean Speech from Noisy Audio: using this module will upload noise audio and then propose algorithm will clean audio and then generate new audio file called ‘clean.wav’.

SCREEN SHOTS

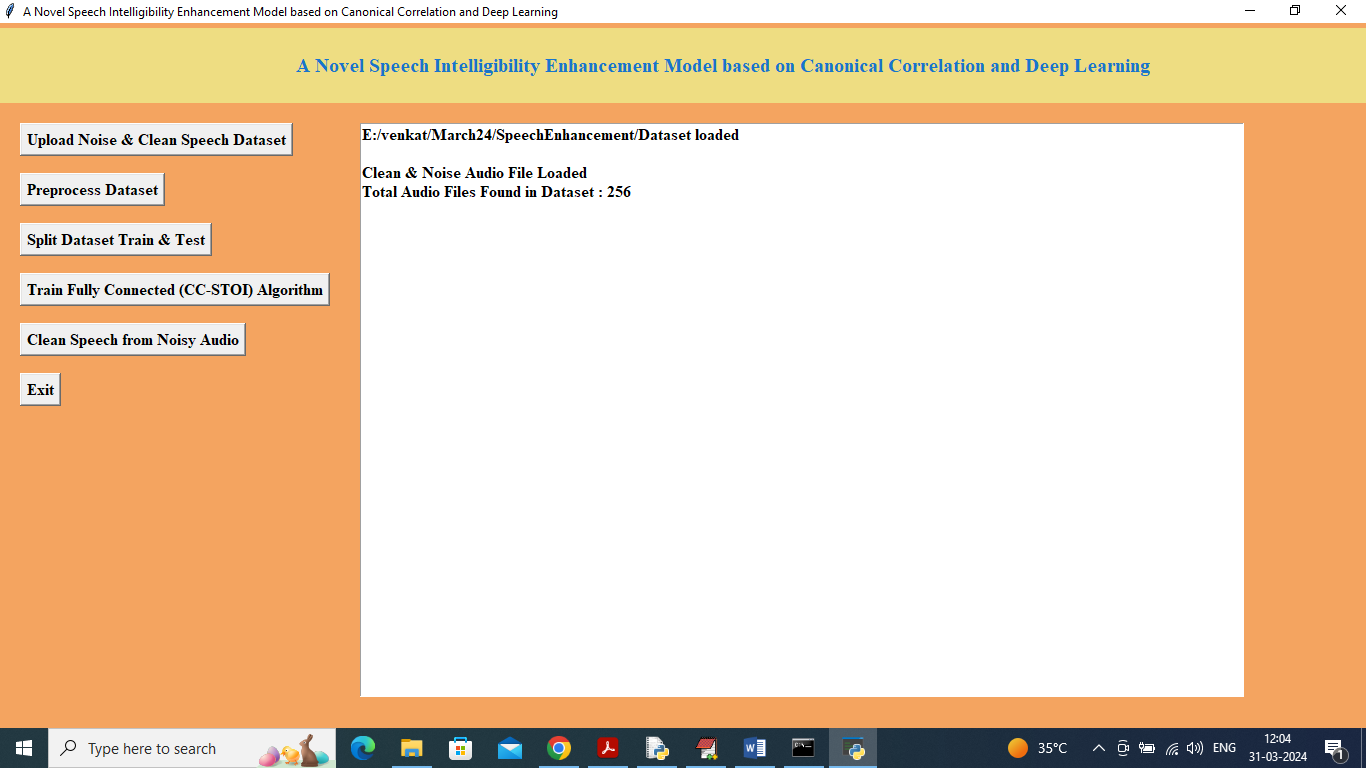
To run project double click on ‘run.bat’ file to get below screen



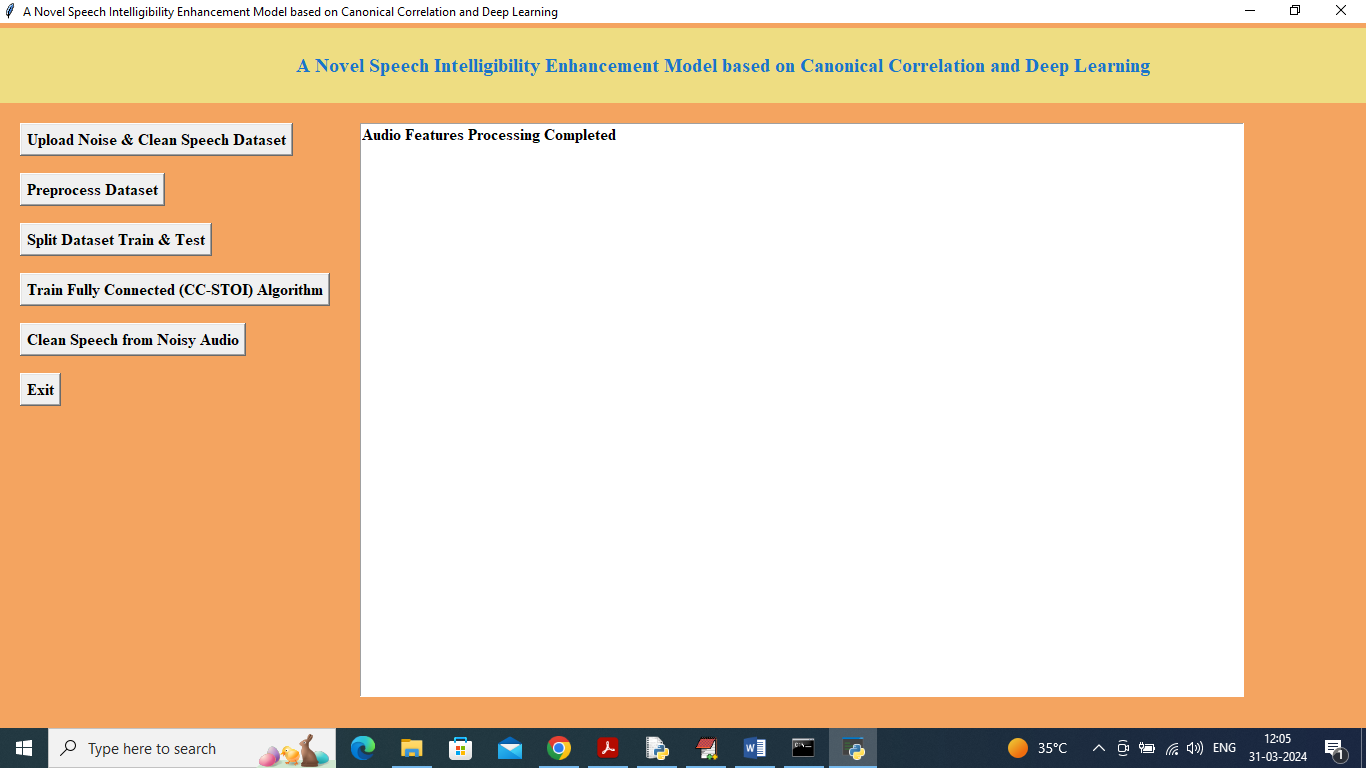
In above screen click on ‘Upload Noise & Clean Speech Dataset’ button to upload dataset and get below page



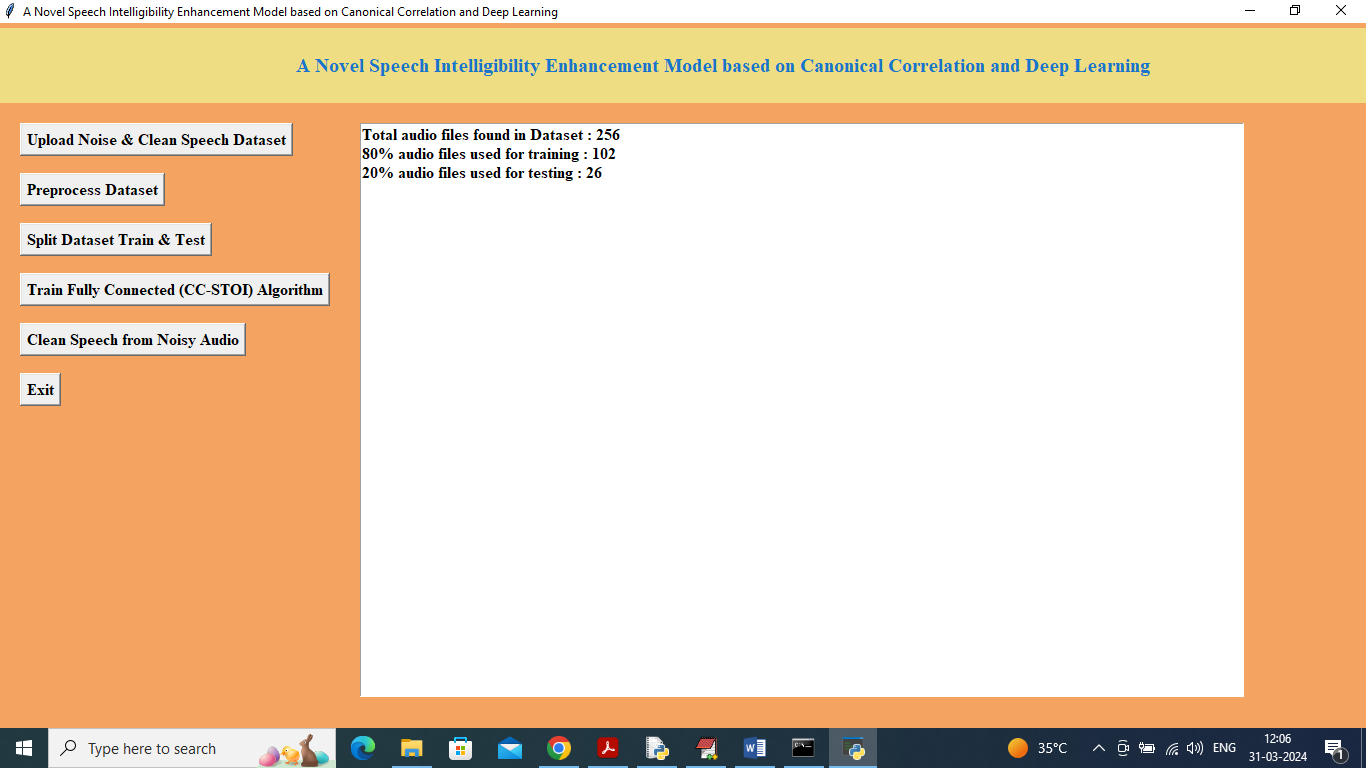
In above screen selecting and uploading ‘Dataset’ folder to upload dataset and get below page



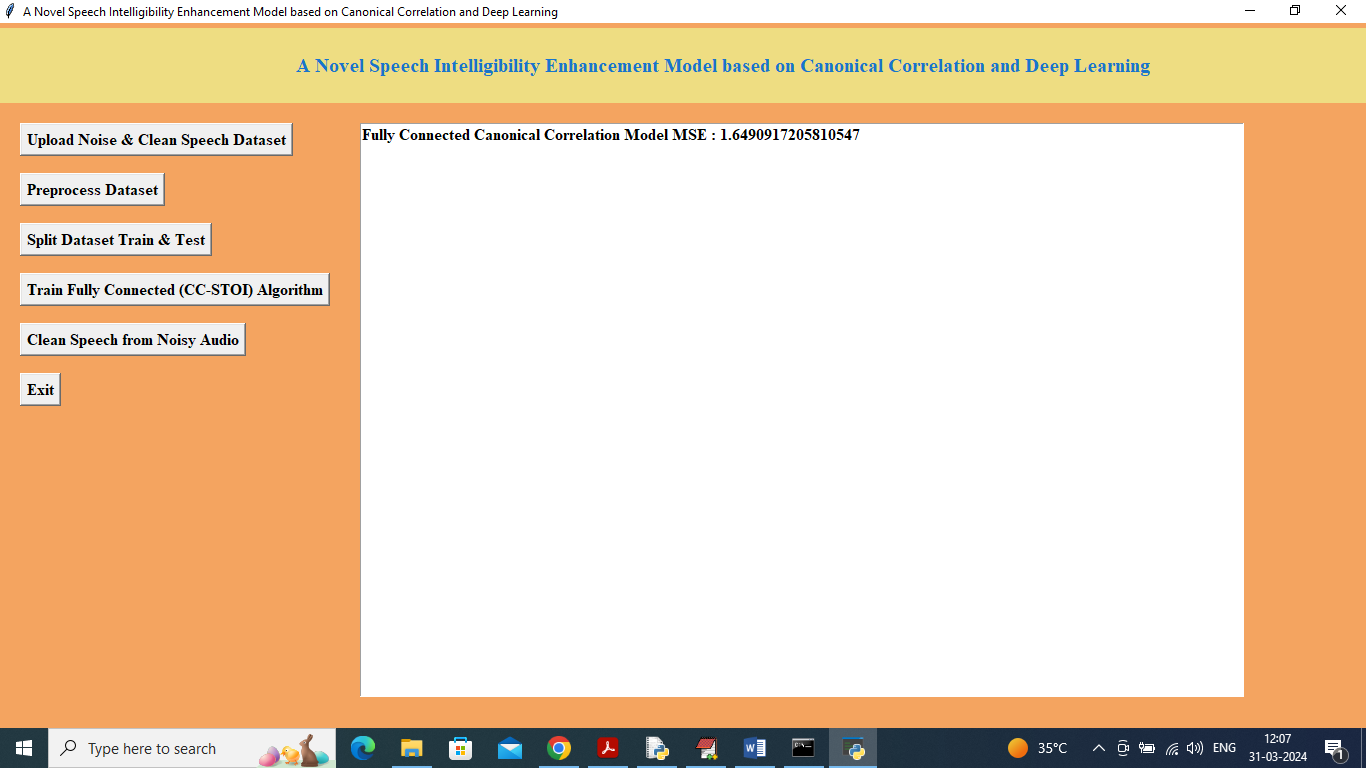
In above screen dataset loaded and can see total number of audio files found in dataset and now click on ‘Pre-process Dataset’ button to clean and normalize audio features and get below page



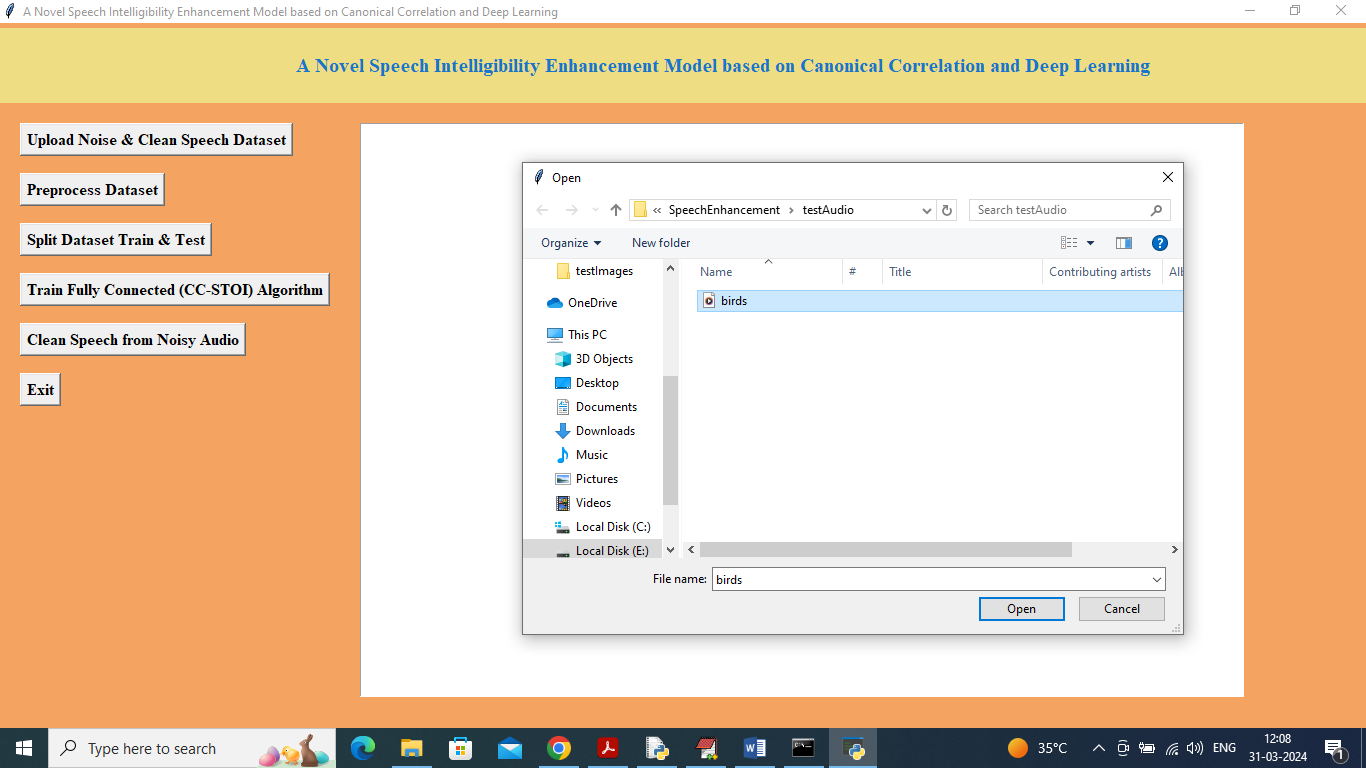
In above screen dataset processing completed and now click on ‘Split Dataset Train & Test’ button to split dataset and get below page



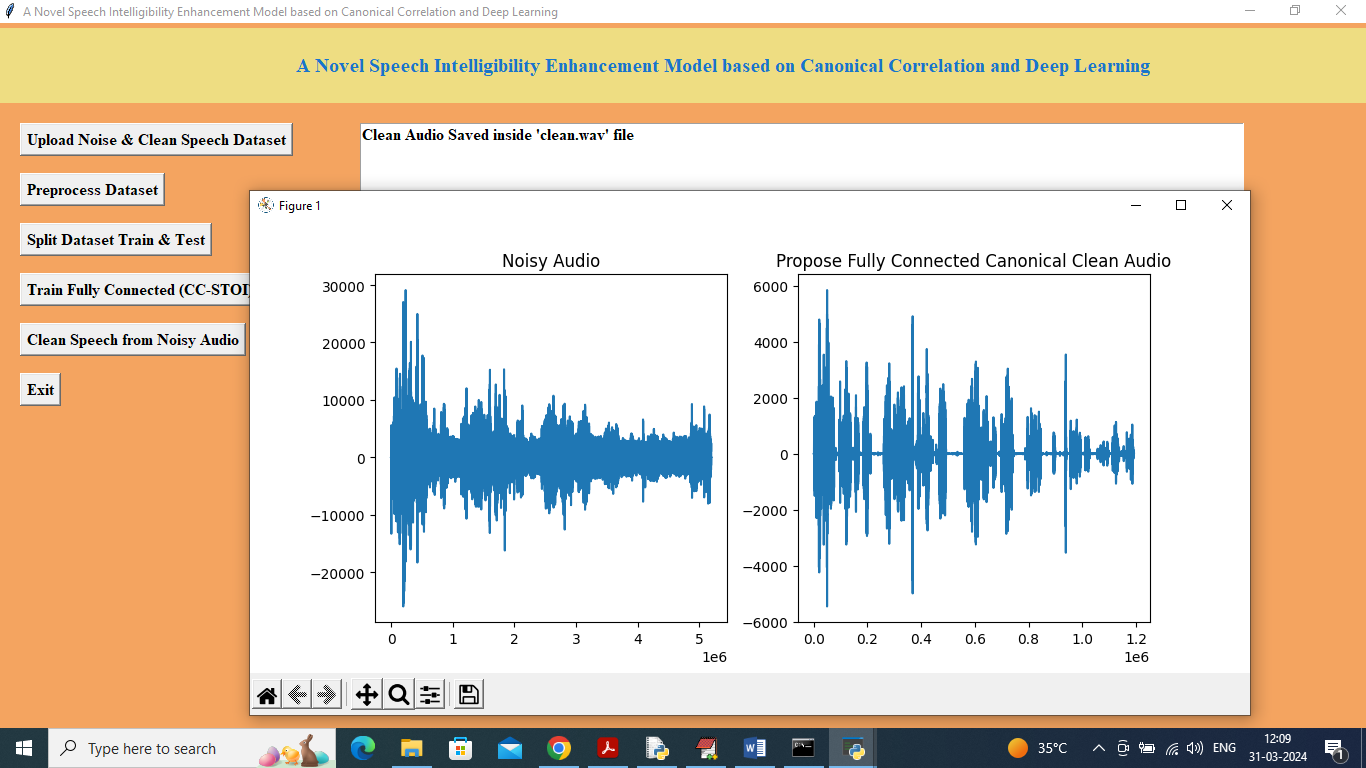
In above screen both clean and noise audio files are 256 which mean noise 128 and clean 128 so application will use training noise 80% as training data and 20% as testing data and now click on ‘Train Fully Connected (CC-STOI) Algorithm’ button to train model and get below page



In above screen propose algorithm got 1.64 as MSE score and now click on ‘Clean Speech from Noisy Audio’ button to upload test audio file and get below output



In above screen selecting and uploading ‘birds.wav’ noise audio file and then click on ‘Open’ button to get below output



In above screen in text area can see clean audio file saved in ‘clean.wav’ file and in graph where first graph contains original audio file signals and can see there are more noise (blue lines refers to noise) and in second graph we can see clean predicted audio file which has less blue lines which means noise is removed and speech is enhanced.

Clean audio you can hear by double click on clean.wav file

Similarly by following above screens you can clean any noise audio file