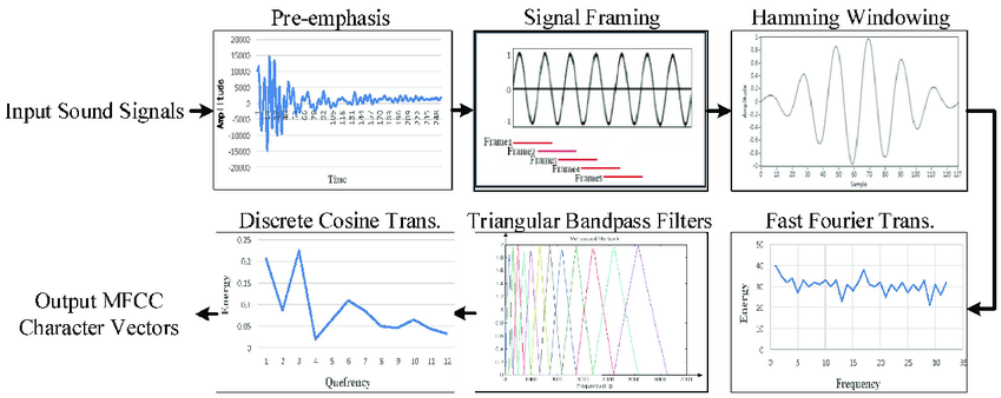
Question 1 :

In sound processing, the Mel-frequency cepstrum (MFC) is a representation of a sound's short-term power spectrum. It is derived through a linear cosine transform of the log power spectrum, calculated on a nonlinear Mel scale of frequency. Mel-frequency cepstral coefficients (MFCCs) are the individual coefficients constituting the MFC representation.

MFCC is a feature extraction technique widely used in speech and audio processing. MFCCs are used to represent the spectral characteristics of sound in a way that is well-suited for various machine learning tasks, such as speech recognition and music analysis. Below figure shows the step-by-step process of extracting MFCC features.



Your task is to predict the Genus of frogs given a set of MFCC features. The dataset for this problem is "Q1Data.csv". The dataset contains 21 MFCC features and its corresponding Family, Genus and Species in the Frog Family.

Part 1 : Decision Trees

a. Write a code **from scratch** to train a Decision Tree Classifier to predict the Genus using the **first 6 features (MFCC_1 ,…., MFCC_6)** of given dataset by splitting the dataset into train and test in the ratio of **20:80**.
b. Plot the test accuracy by pruning the tree to a **depth ranging from 1,2,…..15.**
c. What are your observations from the graph plotted above.

Part 2 : k Nearest Neighbours

a. Write a code **from scratch** which predicts the Genus using the kNN classifier **using all the 21 features**.
b. Extract **30 % of the dataset for the reference/training dataset** and **10% as test dataset**. You do not require to use the entire dataset for this question. Plot the test accuracy for **k ranging from 1,2…..,50.**
c. Report your observations from the above graph.

Part 3 : Using Scikit-Learn functions

a. Now use the inbuilt functions of sklearn library for both Decision Trees and k Nearest Neighbours.
b. Here you need to use all the features and entire dataset.
    i. For Decision Tree : **Train Test ratio is 80:20** and **depth varies from 1,…15**
    ii. For kNN : **Train Test ratio is 80:20** and **k varies from 1,….,1000.**
c. Plot the test accuracies as you have done in Part 1 and 2.
d. Plot all 4 graphs together on a subplot and report your observations comparing the accuracy of model built from Scratch and inbuilt function for both the classifiers.