Abstract

Speech is one of the most innate ways the human express themselves. We recognize its significance while using other communication like emails and text messages where we frequently utilize emotion to convey our feelings. Emotion detection has become an essential in today's digital world of remote communication. The speech which is in fear, sad, joy have higher and wide range in pitch whereas have low range in pitch. In existing system SVM and MLP techniques are used to predict the speech emotion. Speech based emotion recognition (SER) system where different emotions are recognized by means of convolution neural network (CNN) and recurrent neural network (RNN) classifiers. Librosa package in python language is used to develop proposed algorithm and its performance is tested on Crema Audio dataset of emotional speech (CREMA) dataset to differentiate emotions such as happiness, surprise, anger, sad, fear, etc. The output makes the calculations based on the fundamental frequency of each speech frame to map the raw speech data straight to a textured image. Using deep neural network models for emotion recognition, the textured images produced by the conversion can be categorized.