

# Spoken Digit Recognition

In this notebook, You will do Spoken Digit Recognition.

Input - speech signal, output - digit number

It contains

1. Reading the dataset. and Preprocess the data set. Detailed instructions are given below. You have to write the code in the same cell which contains the instruction.
2. Training the LSTM with RAW data
3. Converting to spectrogram and Training the LSTM network
4. Creating the augmented data and doing step 2 and 3 again.

## Instructions:

1. Don't change any Grader Functions. Don't manipulate any Grader functions. If you manipulate any, it will be considered as plagiarised.
2. Please read the instructions on the code cells and markdown cells. We will explain what to write.
3. Please return outputs in the same format what we asked. Eg. Don't return List of we are asking for a numpy array.
4. Please read the external links that we are given so that you will learn the concept behind the code that you are writing.
5. We are giving instructions at each section if necessary, please follow them.

## Every Grader function has to return True.

```
In [ ]: import numpy as np
import pandas as pd
import librosa
import os
##if you need any imports you can do that here.
```

```
In [ ]: !gdown --id 17YGQheavMbM6aeHYjUcGssXfb7eQH01z
```

```
/usr/local/lib/python3.7/dist-packages/gdown/cli.py:131: FutureWarning: Option `--id` was deprecated in version 4.3.1 and will be removed in 5.0. You don't need to pass it anymore to use a file ID.
  category=FutureWarning,
Downloading...
From: https://drive.google.com/uc?id=17YGQheavMbM6aeHYjUcGssXfb7eQH01z
To: /content/recordings.zip
100% 9.28M/9.28M [00:00<00:00, 250MB/s]
```

```
In [ ]: !unzip "/content/recordings.zip"
```

```
Archive: /content/recordings.zip
creating: recordings/
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

inflating: recordings/5\_nicolas\_14.wav  
inflating: recordings/5\_nicolas\_15.wav  
inflating: recordings/5\_nicolas\_16.wav  
inflating: recordings/5\_nicolas\_17.wav  
inflating: recordings/5\_nicolas\_18.wav  
inflating: recordings/5\_nicolas\_19.wav  
inflating: recordings/5\_nicolas\_20.wav  
inflating: recordings/5\_nicolas\_21.wav  
inflating: recordings/5\_nicolas\_22.wav  
inflating: recordings/5\_nicolas\_23.wav  
inflating: recordings/5\_nicolas\_24.wav  
inflating: recordings/5\_nicolas\_25.wav  
inflating: recordings/5\_nicolas\_26.wav  
inflating: recordings/5\_nicolas\_27.wav  
inflating: recordings/5\_nicolas\_28.wav  
inflating: recordings/5\_nicolas\_29.wav  
inflating: recordings/5\_nicolas\_30.wav  
inflating: recordings/5\_nicolas\_31.wav  
inflating: recordings/5\_nicolas\_32.wav  
inflating: recordings/5\_nicolas\_33.wav  
inflating: recordings/5\_nicolas\_34.wav  
inflating: recordings/5\_nicolas\_35.wav  
inflating: recordings/5\_nicolas\_36.wav  
inflating: recordings/5\_nicolas\_37.wav  
inflating: recordings/5\_nicolas\_38.wav  
inflating: recordings/5\_nicolas\_39.wav  
inflating: recordings/5\_nicolas\_40.wav  
inflating: recordings/5\_nicolas\_41.wav  
inflating: recordings/5\_nicolas\_42.wav  
inflating: recordings/5\_nicolas\_43.wav  
inflating: recordings/5\_nicolas\_44.wav  
inflating: recordings/5\_nicolas\_45.wav  
inflating: recordings/5\_nicolas\_46.wav  
inflating: recordings/5\_nicolas\_47.wav  
inflating: recordings/5\_nicolas\_48.wav  
inflating: recordings/5\_nicolas\_49.wav  
inflating: recordings/5\_theo\_0.wav  
inflating: recordings/5\_theo\_1.wav  
inflating: recordings/5\_theo\_10.wav  
inflating: recordings/5\_theo\_11.wav  
inflating: recordings/5\_theo\_12.wav  
inflating: recordings/5\_theo\_13.wav  
inflating: recordings/5\_theo\_14.wav  
inflating: recordings/5\_theo\_15.wav  
inflating: recordings/5\_theo\_16.wav  
inflating: recordings/5\_theo\_17.wav  
inflating: recordings/5\_theo\_18.wav  
inflating: recordings/5\_theo\_19.wav  
inflating: recordings/5\_theo\_20.wav  
inflating: recordings/5\_theo\_21.wav  
inflating: recordings/5\_theo\_22.wav  
inflating: recordings/5\_theo\_23.wav  
inflating: recordings/5\_theo\_24.wav  
inflating: recordings/5\_theo\_25.wav  
inflating: recordings/5\_theo\_26.wav  
inflating: recordings/5\_theo\_27.wav  
inflating: recordings/5\_theo\_28.wav

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
inflating: recordings/9_jackson_8.wav  
inflating: recordings/9_jackson_9.wav  
inflating: recordings/9_nicolas_0.wav  
inflating: recordings/9_nicolas_1.wav  
inflating: recordings/9_nicolas_10.wav  
inflating: recordings/9_nicolas_11.wav  
inflating: recordings/9_nicolas_12.wav  
inflating: recordings/9_nicolas_13.wav  
inflating: recordings/9_nicolas_14.wav  
inflating: recordings/9_nicolas_15.wav  
inflating: recordings/9_nicolas_16.wav  
inflating: recordings/9_nicolas_17.wa  
inflating: recordings/9_nicolas_18.wav  
inflating: recordings/9_nicolas_19.wav  
inflating: recordings/9_nicolas_2.wav  
inflating: recordings/9_nicolas_20.wav  
inflating: recordings/9_nicolas_21.wav  
inflating: recordings/9_nicolas_22.wav  
inflating: recordings/9_nicolas_23.wav  
inflating: recordings/9_nicolas_24.wav  
inflating: recordings/9_nicolas_25.wav  
inflating: recordings/9_nicolas_26.wav  
inflating: recordings/9_nicolas_27.wav  
inflating: recordings/9_nicolas_28.wav  
inflating: recordings/9_nicolas_29.wav  
inflating: recordings/9_nicolas_3.wav  
inflating: recordings/9_nicolas_30.wav  
inflating: recordings/9_nicolas_31.wav  
inflating: recordings/9_nicolas_32.wav  
inflating: recordings/9_nicolas_33.wav  
inflating: recordings/9_nicolas_34.wav  
inflating: recordings/9_nicolas_35.wav  
inflating: recordings/9_nicolas_36.wav  
inflating: recordings/9_nicolas_37.wav  
inflating: recordings/9_nicolas_38.wav  
inflating: recordings/9_nicolas_39.wav  
inflating: recordings/9_nicolas_4.wav  
inflating: recordings/9_nicolas_40.wav  
inflating: recordings/9_nicolas_41.wav  
inflating: recordings/9_nicolas_42.wav  
inflating: recordings/9_nicolas_43.wav  
inflating: recordings/9_nicolas_44.wav  
inflating: recordings/9_nicolas_45.wav  
inflating: recordings/9_nicolas_46.wav  
inflating: recordings/9_nicolas_47.wav  
inflating: recordings/9_nicolas_48.wav  
inflating: recordings/9_nicolas_49.wav  
inflating: recordings/9_nicolas_5.wav  
inflating: recordings/9_nicolas_6.wav  
inflating: recordings/9_nicolas_7.wav  
inflating: recordings/9_nicolas_8.wav  
inflating: recordings/9_nicolas_9.wav  
inflating: recordings/9_theo_0.wav  
inflating: recordings/9_theo_1.wav  
inflating: recordings/9_theo_10.wav  
inflating: recordings/9_theo_11.wav  
inflating: recordings/9_theo_12.wav  
inflating: recordings/9_theo_13.wav  
inflating: recordings/9_theo_14.wav  
inflating: recordings/9_theo_15.wav  
inflating: recordings/9_theo_16.wav  
inflating: recordings/9_theo_17.wav  
inflating: recordings/9_theo_18.wav  
inflating: recordings/9_theo_19.wav  
inflating: recordings/9_theo_2.wav  
inflating: recordings/9_theo_20.wav
```

[illegible]

```

inflating: recordings/9_yweweler_36.wav
inflating: recordings/9_yweweler_37.wav
inflating: recordings/9_yweweler_38.wav
inflating: recordings/9_yweweler_39.wav
inflating: recordings/9_yweweler_4.wav
inflating: recordings/9_yweweler_40.wav
inflating: recordings/9_yweweler_41.wav
inflating: recordings/9_yweweler_42.wav
inflating: recordings/9_yweweler_43.wav
inflating: recordings/9_yweweler_44.wav
inflating: recordings/9_yweweler_45.wav
inflating: recordings/9_yweweler_46.wav
inflating: recordings/9_yweweler_47.wav
inflating: recordings/9_yweweler_48.wav
inflating: recordings/9_yweweler_49.wav
inflating: recordings/9_yweweler_5.wav
inflating: recordings/9_yweweler_6.wav
inflating: recordings/9_yweweler_7.wav
inflating: recordings/9_yweweler_8.wav
inflating: recordings/9_yweweler_9.wav

```

We shared recordings.zip, please unzip those.

```

In [ ]: #read the all file names in the recordings folder given by us
        #(if you get entire path, it is very useful in future)
        #save those files names as list in "all_files"
        all_files = os.listdir("/content/recordings/")
        #getting labels:
        labels = [f[0] for f in all_files]
        all_files = ['/content/recordings/'+f for f in all_files]

        print(all_files[:5])

```

```

['/content/recordings/0_jackson_29.wav', '/content/recordings/2_yweweler_46.wav', '/content/recordings/0_nicolas_15.wav', '/content/recordings/9_yweweler_12.wav', '/content/recordings/9_jackson_39.wav']

```

## Grader function 1

```

In [ ]: def grader_files():
        temp = len(all_files)==2000
        temp1 = all([x[-3:]=="wav" for x in all_files])
        temp = temp and temp1
        return temp
        grader_files()

```

Out[ ]: True

Create a dataframe(name=df\_audio) with two columns(path, label).

You can get the label from the first letter of name.

Eg: 0\_jackson\_0 --> 0

0\_jackson\_43 --> 0

## Exploring the sound dataset

```

In [ ]: #It is a good programming practise to explore the dataset that you are dealing with. This
        #https://colab.research.google.com/github/Tyler-Hilbert/AudioProcessingInPythonWorkshop/
        #visualize the data and write code to play 2-3 sound samples in the notebook for better
        #please go through the following reference video https://www.youtube.com/watch?v=37zCgCd

```



```
In [ ]: from scipy.io import wavfile

sampling_rate, samples = wavfile.read(f'/content/recordings/0_jackson_29.wav')
```

```
In [ ]: sampling_rate
```

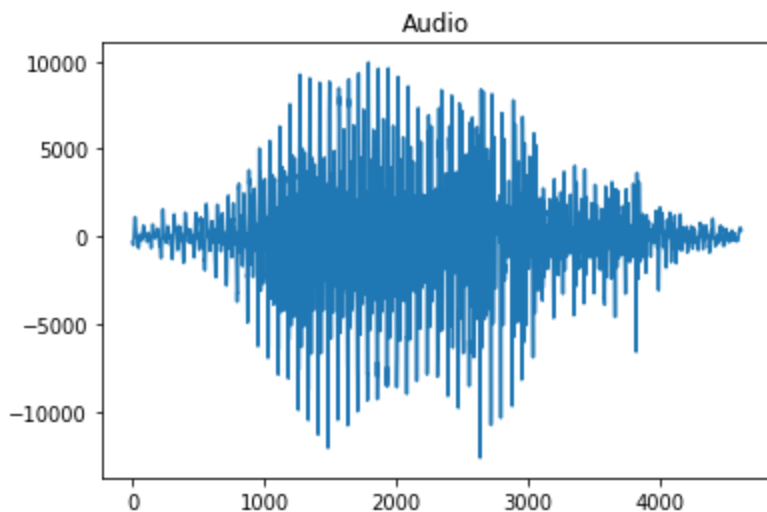
```
Out[ ]: 8000
```

```
In [ ]: samples
```

```
Out[ ]: array([-317, -346, -377, ..., 451, 432, 351], dtype=int16)
```

```
In [ ]: import matplotlib.pyplot as plt
```

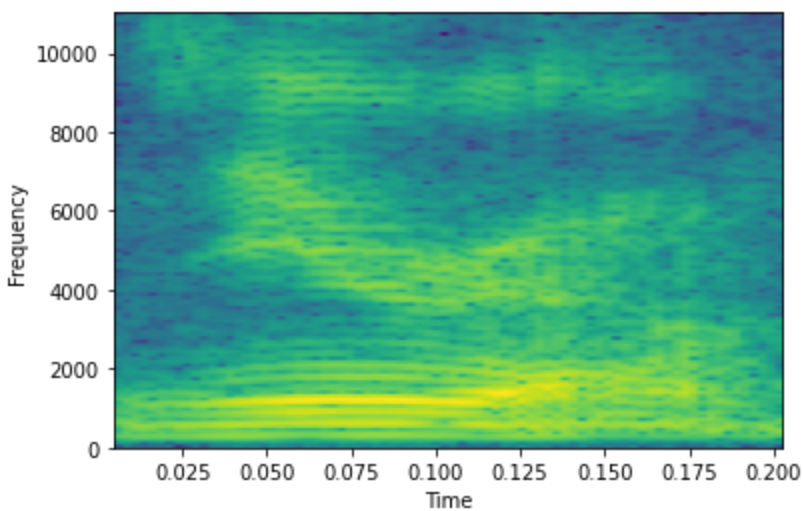
```
plt.plot(samples)
plt.title('Audio')
plt.show()
```



```
In [ ]: import wave
import numpy as np

signal_wave = wave.open('/content/recordings/0_jackson_29.wav', 'r')
sample_rate = 22050
sig = np.frombuffer(signal_wave.readframes(sample_rate), dtype=np.int16)
plt.specgram(sig, Fs=sample_rate, noverlap = 200)
plt.xlabel('Time')
plt.ylabel('Frequency')

plt.show()
```



## Creating dataframe

```
In [ ]: #Create a dataframe(name=df_audio) with two columns(path, label).
        #You can get the label from the first letter of name.
        #Eg: 0_jackson_0 --> 0
        #0_jackson_43 --> 0
        df_audio = pd.DataFrame({'path': all_files, 'label': labels})
        df_audio.head()
```

```
Out[ ]:
```

	path	label
0	/content/recordings/0_jackson_29.wav	0
1	/content/recordings/2_yweweler_46.wav	2
2	/content/recordings/0_nicolas_15.wav	0
3	/content/recordings/9_yweweler_12.wav	9
4	/content/recordings/9_jackson_39.wav	9

```
In [ ]: #info
        df_audio.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype
---  ---
 0   path        2000 non-null   object
 1   label       2000 non-null   object
dtypes: object(2)
memory usage: 31.4+ KB
```

## Grader function 2

```
In [ ]: def grader_df():
        flag_shape = df_audio.shape==(2000,2)
        flag_columns = all(df_audio.columns==['path', 'label'])
        list_values = list(df_audio.label.value_counts())
        flag_label = len(list_values)==10
        flag_label2 = all([i==200 for i in list_values])
        final_flag = flag_shape and flag_columns and flag_label and flag_label2
        return final_flag
        grader_df()
```

Out[ ]: True

```
In [ ]: from sklearn.utils import shuffle
df_audio = shuffle(df_audio, random_state=33) #don't change the random state
```

## Train and Validation split

```
In [ ]: #split the data into train and validation and save in X_train, X_test, y_train, y_test
#use stratify sampling
#use random state of 45
#use test size of 30%
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(df_audio['path'], df_audio['label'],
                                                  stratify = df_audio['label'], random_
```

## Grader function 3

```
In [ ]: def grader_split():
    flag_len = (len(X_train)==1400) and (len(X_test)==600) and (len(y_train)==1400) and
    values_ytrain = list(y_train.value_counts())
    flag_ytrain = (len(values_ytrain)==10) and (all([i==140 for i in values_ytrain]))
    values_ytest = list(y_test.value_counts())
    flag_ytest = (len(values_ytest)==10) and (all([i==60 for i in values_ytest]))
    final_flag = flag_len and flag_ytrain and flag_ytest
    return final_flag
grader_split()
```

Out[ ]: True

## Preprocessing

All files are in the "WAV" format. We will read those raw data files using the librosa

```
In [ ]: sample_rate = 22050
def load_wav(x, get_duration=True):
    '''This return the array values of audio with sampling rate of 22050 and Duration'''
    #loading the wav file with sampling rate of 22050
    samples, sample_rate = librosa.load(x, sr=22050)
    if get_duration:
        duration = librosa.get_duration(samples, sample_rate)
        return [samples, duration]
    else:
        return samples
```

```
In [ ]: #use load_wav function that was written above to get every wave.
#save it in X_train_processed and X_test_processed
# X_train_processed/X_test_processed should be dataframes with two columns(raw_data, dur

raw_train = []
raw_test = []
dur_train = []
dur_test = []

for x in X_train.values:
    r, d = load_wav(x)
    raw_train.append(r)
```

```

dur_train.append(d)

for x in X_test.values:
    r, d = load_wav(x)
    raw_test.append(r)
    dur_test.append(d)

X_train_processed = pd.DataFrame({'raw_data' : raw_train, 'duration' : dur_train})
X_test_processed = pd.DataFrame({'raw_data' : raw_test, 'duration' : dur_test})

```

```

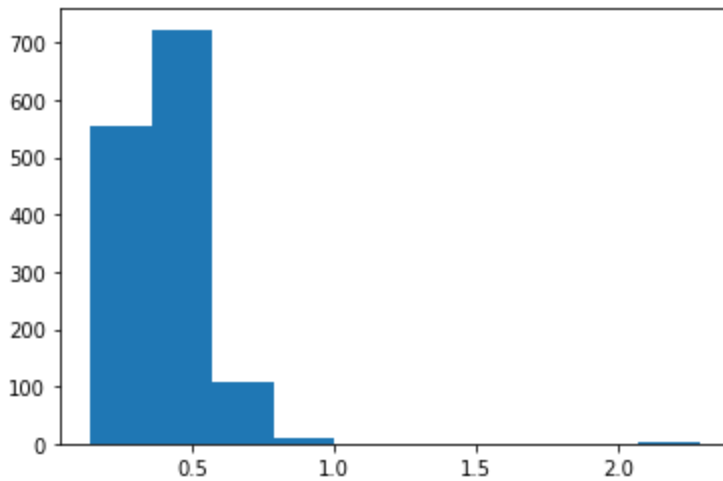
In [ ]: #plot the histogram of the duration for train
plt.hist(list(X_train_processed['duration']))

```

```

Out[ ]: (array([556., 723., 107., 11., 1., 0., 0., 0., 0., 2.]),
 array([0.14353741, 0.35746032, 0.57138322, 0.78530612, 0.99922902,
        1.21315193, 1.42707483, 1.64099773, 1.85492063, 2.06884354,
        2.28276644]),
 <a list of 10 Patch objects>)

```



```

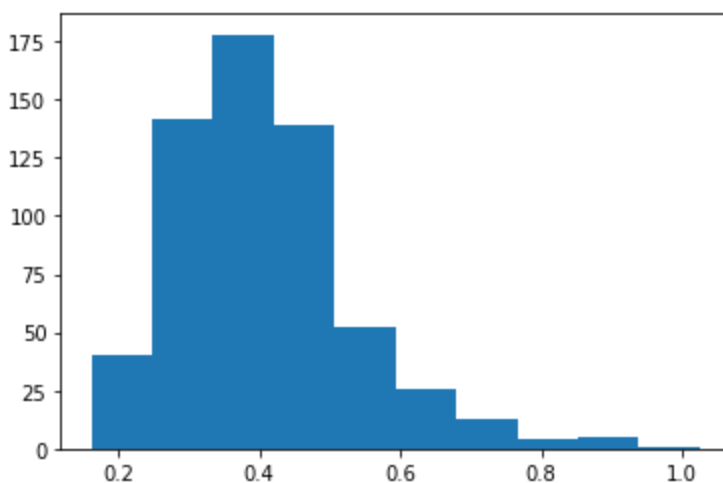
In [ ]: #plot the histogram of the duration for test
plt.hist(list(X_test_processed['duration']))

```

```

Out[ ]: (array([ 40., 142., 178., 139., 52., 26., 13., 4., 5., 1.]),
 array([0.16104308, 0.24745578, 0.33386848, 0.42028118, 0.50669388,
        0.59310658, 0.67951927, 0.76593197, 0.85234467, 0.93875737,
        1.02517007]),
 <a list of 10 Patch objects>)

```



```

In [ ]: #print 0 to 100 percentile values with step size of 10 for train data duration.
for i in range(0,110, 10):
    print(i, "th percentile value:", X_train_processed['duration'].quantile(i/100))

```

```

0 th percentile value: 0.1435374149659864
10 th percentile value: 0.25874829931972787

```

```

20 th percentile value: 0.29872108843537415
30 th percentile value: 0.33165532879818593
40 th percentile value: 0.35830385487528343
50 th percentile value: 0.38988662131519275
60 th percentile value: 0.4158730158730159
70 th percentile value: 0.44721995464852604
80 th percentile value: 0.48358276643990933
90 th percentile value: 0.553201814058957
100 th percentile value: 2.282766439909297

```

```

In [ ]: ##print 90 to 100 percentile values with step size of 1.
        for i in range(90,101, 1):
            print(i, "th percentile value:", X_train_processed['duration'].quantile(i/100))

90 th percentile value: 0.553201814058957
91 th percentile value: 0.5669832199546486
92 th percentile value: 0.5778467120181406
93 th percentile value: 0.5915165532879821
94 th percentile value: 0.6082149659863945
95 th percentile value: 0.6203106575963718
96 th percentile value: 0.6332625850340134
97 th percentile value: 0.6528367346938775
98 th percentile value: 0.674312925170068
99 th percentile value: 0.7832281179138321
100 th percentile value: 2.282766439909297

```

## Grader function 4

```

In [ ]: def grader_processed():
        flag_columns = (all(X_train_processed.columns==['raw_data', 'duration'])) and (all(X
        flag_shape = (X_train_processed.shape==(1400, 2)) and (X_test_processed.shape==(600
        return flag_columns and flag_shape
        grader_processed()

```

Out[ ]: True

**Based on our analysis 99 percentile values are less than 0.8sec so we will limit maximum length of X\_train\_processed and X\_test\_processed to 0.8 sec. It is similar to pad\_sequence for a text dataset.**

**While loading the audio files, we are using sampling rate of 22050 so one sec will give array of length 22050. so, our maximum length is  $0.8 \times 22050 = 17640$  Pad with Zero if length of sequence is less than 17640 else Truncate the number.**

**Also create a masking vector for train and test.**

**masking vector value = 1 if it is real value, 0 if it is pad value. Masking vector data type must be bool.**

```

In [ ]: max_length = 17640

```

```

In [ ]: X_train_processed.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1400 entries, 0 to 1399
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  -
0   raw_data    1400 non-null   object
1   duration    1400 non-null   float64
dtypes: float64(1), object(1)
memory usage: 22.0+ KB

```

```

In [ ]: X_test_processed.info()

```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 600 entries, 0 to 599
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  -
0   raw_data    600 non-null    object
1   duration    600 non-null    float64
dtypes: float64(1), object(1)
memory usage: 9.5+ KB
```

```
In [ ]: ## as discussed above, Pad with Zero if length of sequence is less than 17640 else Trunc
## save in the X_train_pad_seq, X_test_pad_seq
## also Create masking vector X_train_mask, X_test_mask

## all the X_train_pad_seq, X_test_pad_seq, X_train_mask, X_test_mask will be numpy arra
from tensorflow.keras.preprocessing.sequence import pad_sequences

X_train_pad_seq = pad_sequences(X_train_processed.raw_data,
                                maxlen = max_length,
                                padding = 'post',
                                dtype = 'float64',
                                truncating = 'post')
X_test_pad_seq = pad_sequences(X_test_processed.raw_data,
                                maxlen = max_length,
                                padding = 'post',
                                dtype = 'float64',
                                truncating = 'post')

X_train_mask = np.array(X_train_pad_seq != 0)
X_test_mask = np.array(X_test_pad_seq != 0)
```

## Grader function 5

```
In [ ]: def grader_padoutput():
    flag_padshape = (X_train_pad_seq.shape==(1400, 17640)) and (X_test_pad_seq.shape==(6
    flag_maskshape = (X_train_mask.shape==(1400, 17640)) and (X_test_mask.shape==(600, 1
    flag_dtype = (X_train_mask.dtype==bool) and (X_test_mask.dtype==bool)
    return flag_padshape and flag_maskshape and flag_dtype
grader_padoutput()
```

Out[ ]: True

```
In [ ]: X_train_pad_seq
```

```
Out[ ]: array([[ 1.72840417e-04,  1.99094706e-04,  1.51869914e-04, ...,
                0.00000000e+00,  0.00000000e+00,  0.00000000e+00],
               [-1.13962009e-03, -1.32962456e-03, -1.35439681e-03, ...,
                0.00000000e+00,  0.00000000e+00,  0.00000000e+00],
               [-3.40759940e-03, -5.44826675e-04,  1.95389055e-03, ...,
                0.00000000e+00,  0.00000000e+00,  0.00000000e+00],
               ...,
               [-1.38692689e-04, -3.24455381e-04, -4.80764866e-04, ...,
                0.00000000e+00,  0.00000000e+00,  0.00000000e+00],
               [-1.25074387e-03,  5.40220644e-04,  1.74343493e-03, ...,
                0.00000000e+00,  0.00000000e+00,  0.00000000e+00],
               [-2.2223989e-05,  1.69483188e-04,  2.79781147e-04, ...,
                0.00000000e+00,  0.00000000e+00,  0.00000000e+00]])
```

```
In [ ]: X_train_pad_seq.shape
```

Out[ ]: (1400, 17640)

```
In [ ]: np.expand_dims(X_train_pad_seq, -1)
```

```

Out[ ]: array([[ 1.72840417e-04],
           [ 1.99094706e-04],
           [ 1.51869914e-04],
           ...,
           [ 0.00000000e+00],
           [ 0.00000000e+00],
           [ 0.00000000e+00]],

          [[-1.13962009e-03],
           [-1.32962456e-03],
           [-1.35439681e-03],
           ...,
           [ 0.00000000e+00],
           [ 0.00000000e+00],
           [ 0.00000000e+00]],

          [[-3.40759940e-03],
           [-5.44826675e-04],
           [ 1.95389055e-03],
           ...,
           [ 0.00000000e+00],
           [ 0.00000000e+00],
           [ 0.00000000e+00]],

          ...,

          [[-1.38692689e-04],
           [-3.24455381e-04],
           [-4.80764866e-04],
           ...,
           [ 0.00000000e+00],
           [ 0.00000000e+00],
           [ 0.00000000e+00]],

          [[-1.25074387e-03],
           [ 5.40220644e-04],
           [ 1.74343493e-03],
           ...,
           [ 0.00000000e+00],
           [ 0.00000000e+00],
           [ 0.00000000e+00]],

          [[-2.22223989e-05],
           [ 1.69483188e-04],
           [ 2.79781147e-04],
           ...,
           [ 0.00000000e+00],
           [ 0.00000000e+00],
           [ 0.00000000e+00]])

```

```

In [ ]: np.expand_dims(X_train_pad_seq, -1).shape

```

```

Out[ ]: (1400, 17640, 1)

```

```

In [ ]: #shape of input adjusted for using LSTM.
X_train_pad_seq = np.expand_dims(X_train_pad_seq, -1)
X_test_pad_seq = np.expand_dims(X_test_pad_seq, -1)

#converting label values in train and test to int
y_train = y_train.values.astype('int')
y_test = y_test.values.astype('int')

```

## 1. Giving Raw data directly.

Now we have

Train data: X\_train\_pad\_seq, X\_train\_mask and y\_train

Test data: X\_test\_pad\_seq, X\_test\_mask and y\_test

We will create a LSTM model which takes this input.

Task:

1. Create an LSTM network which takes "X\_train\_pad\_seq" as input, "X\_train\_mask" as mask input. You can use any number of LSTM cells. Please read LSTM documentation([https://www.tensorflow.org/api\\_docs/python/tf/keras/layers/LSTM](https://www.tensorflow.org/api_docs/python/tf/keras/layers/LSTM)) in tensorflow to know more about mask and also [https://www.tensorflow.org/guide/keras/masking\\_and\\_padding](https://www.tensorflow.org/guide/keras/masking_and_padding)
2. Get the final output of the LSTM and give it to Dense layer of any size and then give it to Dense layer of size 10(because we have 10 outputs) and then compile with the sparse categorical cross entropy(because we are not converting it to one hot vectors). Also check the datatype of class labels(y\_values) and make sure that you convert your class labels to integer datatype before fitting in the model.
3. While defining your model make sure that you pass both the input layer and mask input layer as input to lstm layer as follows

```
lstm_output = self.lstm(input_layer, mask=masking_input_layer)
```

4. Use tensorboard to plot the graphs of loss and metric(use custom micro F1 score as metric) and histograms of gradients. You can write your code for computing F1 score using this [link](#)
5. make sure that it won't overfit.
6. You are free to include any regularization

```
In [ ]: from sklearn.metrics import f1_score
class Metric_f1(tf.keras.callbacks.Callback):
    def __init__(self, x, y, validation_data):
        super().__init__()
        self.train_data = x
        self.y_train = y
        self.test_data = validation_data[0]
        self.y_test = validation_data[1]
        self.history = {}
        self.history['val_f1_score'] = []
        self.history['train_f1_score'] = []

    def on_epoch_end(self, epochs, logs = {}):
        l = logs.get('val_f1_score')
        if l is not None:
            val_f1_score = l[-1]

        train_preds = np.argmax(self.model.predict(self.train_data), axis = -1)
        train_f1_score = f1_score(self.y_train, train_preds, average='micro')
        self.history['train_f1_score'].append(train_f1_score)

        test_preds = np.argmax(self.model.predict(self.test_data), axis = -1)
        test_f1_score = f1_score(y_test, test_preds, average='micro')
        self.history['val_f1_score'].append(test_f1_score)

        print("f1_score:", train_f1_score, " val_f1_score:", test_f1_score)

        writer1 = tf.summary.create_file_writer(log_dir + '/train_f1_score')
        writer2 = tf.summary.create_file_writer(log_dir + '/validation_f1_score')

        with writer1.as_default():
```



```

        tf.summary.scalar('F1_Score', train_f1_score, step=epochs)
        writer1.flush()

    with writer2.as_default():
        tf.summary.scalar('F1_Score', test_f1_score, step=epochs)
        writer2.flush()

```

```

In [ ]: from tensorflow.keras.layers import Input, LSTM, Dense
        from tensorflow.keras.models import Model
        import tensorflow as tf

```

```

In [ ]: ## as discussed above, please write the architecture of the model.
        ## you will have two input layers in your model (data input layer and mask input layer)
        ## make sure that you have defined the data type of masking layer as bool

```

```

In [ ]: input_layer = Input(shape = (17640,1))
        input_mask = Input(shape = (17640), dtype = 'bool')
        lstm_layer = LSTM(50)(input_layer, mask = input_mask)
        dense_layer = Dense(100, activation = 'relu', kernel_initializer = tf.keras.initializers
        output_layer = Dense(10, activation = 'softmax')(dense_layer)

        model1 = Model(inputs = [input_layer, input_mask], outputs = output_layer)

        model1.summary()

```

Model: "model"

Layer (type)	Output Shape	Param #	Connected to
=====			
input_1 (InputLayer)	[(None, 17640, 1)]	0	[]
input_2 (InputLayer)	[(None, 17640)]	0	[]
lstm (LSTM)	(None, 50)	10400	['input_1[0][0]', 'input_2[0][0]']
dense (Dense)	(None, 100)	5100	['lstm[0][0]']
dense_1 (Dense)	(None, 10)	1010	['dense[0][0]']
=====			
Total params: 16,510			
Trainable params: 16,510			
Non-trainable params: 0			

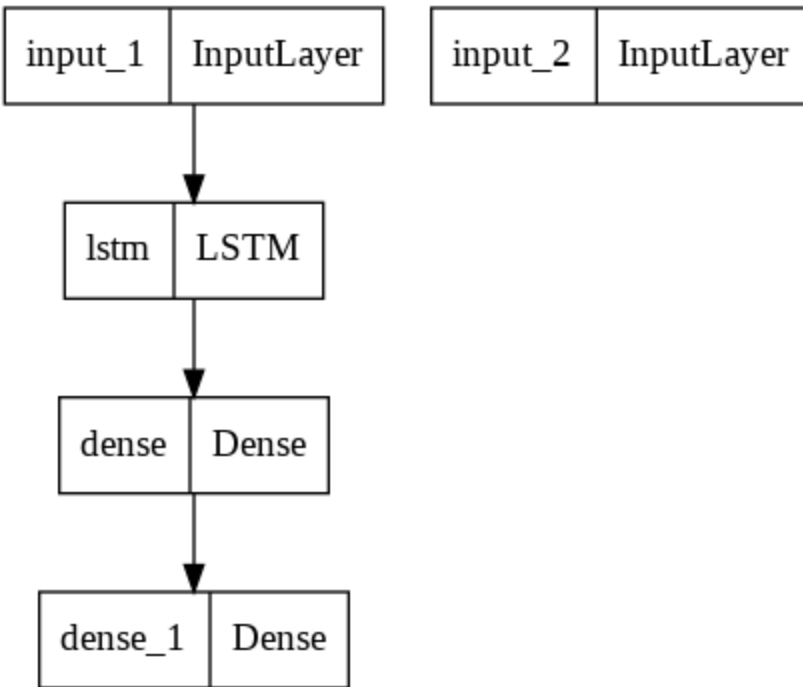
```

In [ ]: tf.keras.utils.plot_model(
        model1, to_file='model_1.png', show_shapes=False, show_layer_names=True,

```

```
rankdir='TB', expand_nested=False, dpi=96
)
```

Out[ ]:



```
In [ ]: model1.compile(optimizer = tf.keras.optimizers.Adam(0.006), loss = 'sparse_categorical_crossentropy')
```

```
In [ ]: tf.keras.backend.clear_session()
!rm -rf ./logs/
```

```
In [ ]: #train your model
#model1.fit([X_train_pad_seq,X_train_mask],y_train_int,.....)

%load_ext tensorboard
import datetime, os
from tensorflow.keras.callbacks import ModelCheckpoint

f1_fn = Metric_f1([X_train_pad_seq,X_train_mask], y_train, ([X_test_pad_seq, X_test_mask], y_test))

log_dir = os.path.join("logs", 'fits', datetime.datetime.now().strftime("%Y%m%d-%H%M%S"))
tensorboard_callback = tf.keras.callbacks.TensorBoard(log_dir = log_dir,histogram_freq=1)
model1.fit(x = [X_train_pad_seq,X_train_mask], y = y_train, epochs = 5, verbose = 1, batch_size = 10,
          validation_data = ([X_test_pad_seq, X_test_mask], y_test) ,
          callbacks = [tensorboard_callback, f1_fn])
```

The tensorboard extension is already loaded. To reload it, use:

```
%reload_ext tensorboard
```

```
Epoch 1/5
10/10 [=====] - ETA: 0s - loss: 2.3069f1_score: 0.10000000000000002
0002 val_f1_score: 0.10000000000000002
10/10 [=====] - 30s 3s/step - loss: 2.3069 - val_loss: 2.3029
Epoch 2/5
10/10 [=====] - ETA: 0s - loss: 2.3045f1_score: 0.10000000000000002
0002 val_f1_score: 0.10000000000000002
10/10 [=====] - 23s 3s/step - loss: 2.3045 - val_loss: 2.3029
Epoch 3/5
10/10 [=====] - ETA: 0s - loss: 2.3039f1_score: 0.10071428571428571
8571 val_f1_score: 0.10166666666666667
10/10 [=====] - 23s 2s/step - loss: 2.3039 - val_loss: 2.3028
Epoch 4/5
10/10 [=====] - ETA: 0s - loss: 2.3031f1_score: 0.10071428571428571
8571 val_f1_score: 0.10166666666666667
10/10 [=====] - 23s 3s/step - loss: 2.3031 - val_loss: 2.3026
Epoch 5/5
```

```

10/10 [=====] - ETA: 0s - loss: 2.3033f1_score: 0.1000000000000000
0002 val_f1_score: 0.10666666666666669
10/10 [=====] - 23s 3s/step - loss: 2.3033 - val_loss: 2.3026
Out[ ]: <keras.callbacks.History at 0x7faa08446d90>

```

```

In [ ]: %tensorboard --logdir logs/fits

```

```

In [ ]: f1_fn.history

```

```

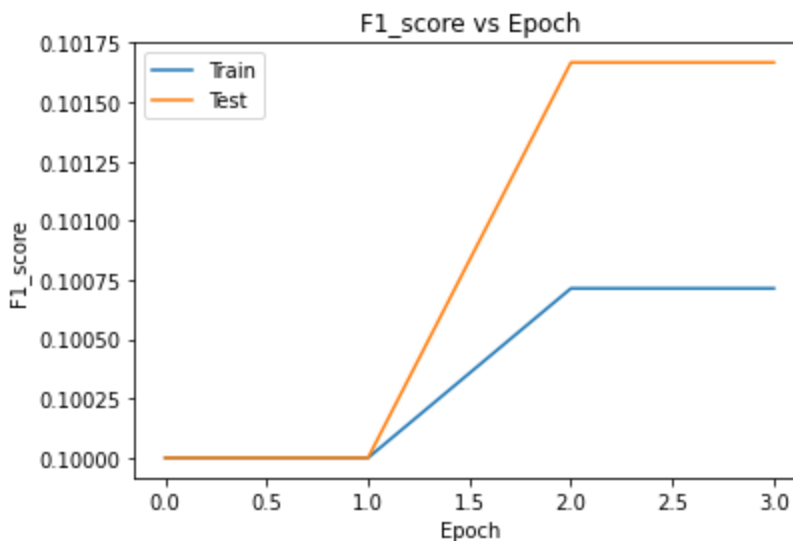
Out[ ]: {'val_f1_score': [0.10000000000000002,
 0.10000000000000002,
 0.10166666666666667,
 0.10166666666666667,
 0.10666666666666669],
'train_f1_score': [0.10000000000000002,
 0.10000000000000002,
 0.10071428571428571,
 0.10071428571428571,
 0.10000000000000002]}

```

```

In [ ]: plt.plot(f1_fn.history['train_f1_score'][:4])
plt.plot(f1_fn.history['val_f1_score'][:4])
plt.title("F1_score vs Epoch")
plt.xlabel('Epoch')
plt.ylabel('F1_score')
plt.legend(["Train", "Test"])
plt.show()

```



## 2. Converting into spectrogram and giving spectrogram data as input

We can use librosa to convert raw data into spectrogram. A spectrogram shows the features in a two-dimensional representation with the intensity of a frequency at a point in time i.e we are converting Time domain to frequency domain. you can read more about this in <https://pnsn.org/spectrograms/what-is-a-spectrogram>

```

In [ ]: #reducing the dimensions back -> since it expects 1 dim input
X_train_pad_seq = np.squeeze(X_train_pad_seq, axis = -1)
X_test_pad_seq = np.squeeze(X_test_pad_seq, axis = -1)

```

```

In [ ]: def convert_to_spectrogram(raw_data):
    '''converting to spectrogram'''
    spectrum = librosa.feature.melspectrogram(y=raw_data, sr=sample_rate, n_mels=64)

```

```
logmel_spectrum = librosa.power_to_db(S=spectrum, ref=np.max)
return logmel_spectrum
```

```
In [ ]: ##use convert_to_spectrogram and convert every raw sequence in X_train_pad_seq and X_test
## save those all in the X_train_spectrogram and X_test_spectrogram ( These two arrays m
X_train_spectrogram = []
for x in X_train_pad_seq:
    X_train_spectrogram.append(convert_to_spectrogram(x))

X_train_spectrogram = np.array(X_train_spectrogram)

X_test_spectrogram = []
for x in X_test_pad_seq:
    X_test_spectrogram.append(convert_to_spectrogram(x))

X_test_spectrogram = np.array(X_test_spectrogram)
```

## Grader function 6

```
In [ ]: def grader_spectrogram():
        flag_shape = (X_train_spectrogram.shape==(1400,64, 35)) and (X_test_spectrogram.shap
        return flag_shape
grader_spectrogram()
```

Out[ ]: True

Now we have

Train data: X\_train\_spectrogram and y\_train

Test data: X\_test\_spectrogram and y\_test

We will create a LSTM model which takes this input.

Task:

1. Create an LSTM network which takes "X\_train\_spectrogram" as input and has to return output at every time step.
2. Average the output of every time step and give this to the Dense layer of any size. (ex: Output from LSTM will be (None, time\_steps, features) average the output of every time step i.e, you should get (None,time\_steps) and then pass to dense layer )
3. give the above output to Dense layer of size 10( output layer) and train the network with sparse categorical cross entropy.
4. Use tensorboard to plot the graphs of loss and metric(use custom micro F1 score as metric) and histograms of gradients. You can write your code for computing F1 score using this [link](#)
5. make sure that it won't overfit.
6. You are free to include any regularization

```
In [ ]: # write the architecture of the model
#print model.summary and make sure that it is following point 2 mentioned above
input_layer = Input(shape = (64,35,))
lstm_layer = LSTM(512, return_sequences = True)(input_layer)
x = tf.math.reduce_mean(lstm_layer, axis = -1)
dense_layer = Dense(512, activation = 'relu', kernel_initializer = tf.keras.initializers
output_layer = Dense(10, activation = 'softmax')(dense_layer)

model2 = Model(inputs = input_layer, outputs = output_layer)
#printing the model summary
model2.summary()
```

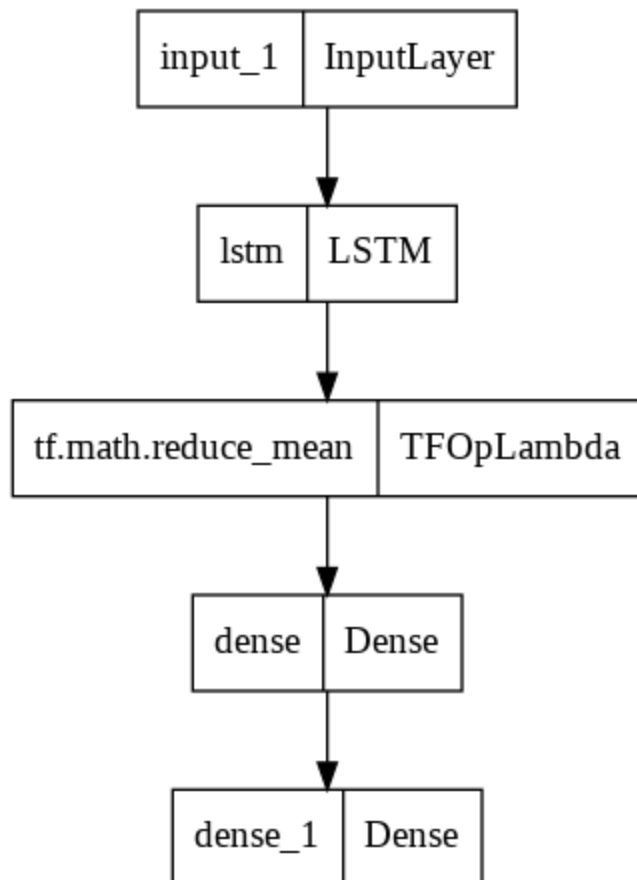
Model: "model"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[ (None, 64, 35) ]	0
lstm (LSTM)	(None, 64, 512)	1122304
tf.math.reduce_mean (TFOpLambda)	(None, 64)	0
dense (Dense)	(None, 512)	33280
dense_1 (Dense)	(None, 10)	5130

=====  
Total params: 1,160,714  
Trainable params: 1,160,714  
Non-trainable params: 0  
=====

```
In [ ]: tf.keras.utils.plot_model(
        model2, to_file='model_2.png', show_shapes=False, show_layer_names=True,
        rankdir='TB', expand_nested=False, dpi=96
    )
```

Out[ ]:



```
In [ ]: #compile and fit your model.
model2.compile(optimizer = tf.keras.optimizers.Adam(0.001), loss = 'sparse_categorical_crossentropy')
#model2.fit([X_train_spectrogram],y_train_int,.....)
```

```
In [ ]: tf.keras.backend.clear_session()
!rm -rf ./logs/
```

```
In [ ]: #train your model
#model1.fit([X_train_pad_seq,X_train_mask],y_train_int,.....)
```

```

%load_ext tensorboard
import datetime, os
from tensorflow.keras.callbacks import ModelCheckpoint

f1_fn = Metric_f1([X_train_spectrogram], y_train, ([X_test_spectrogram], y_test))

log_dir = os.path.join("logs", 'fits', datetime.datetime.now().strftime("%Y%m%d-%H%M%S"))
tensorboard_callback = tf.keras.callbacks.TensorBoard(log_dir = log_dir, histogram_freq=1)
model2.fit(x = [X_train_spectrogram], y = y_train, epochs = 20, verbose = 1, batch_size = 128,
           validation_data = ([X_test_spectrogram], y_test) ,
           callbacks = [tensorboard_callback, f1_fn])

```

The tensorboard extension is already loaded. To reload it, use:

```
%reload_ext tensorboard
```

Epoch 1/20

4/10 [=====>.....] - ETA: 0s - loss: 2.2513

WARNING:tensorflow:Callback method `on\_train\_batch\_end` is slow compared to the batch time (batch time: 0.0039s vs `on\_train\_batch\_end` time: 0.0219s). Check your callbacks.

10/10 [=====] - ETA: 0s - loss: 2.2149f1\_score: 0.3521428571428

5715 val\_f1\_score: 0.36

10/10 [=====] - 1s 130ms/step - loss: 2.2149 - val\_loss: 2.1294

Epoch 2/20

10/10 [=====] - ETA: 0s - loss: 2.0376f1\_score: 0.3307142857142

8574 val\_f1\_score: 0.38666666666666666

10/10 [=====] - 1s 79ms/step - loss: 2.0376 - val\_loss: 1.9525

Epoch 3/20

10/10 [=====] - ETA: 0s - loss: 1.8345f1\_score: 0.44 val\_f1\_sc

ore: 0.41833333333333333

10/10 [=====] - 1s 78ms/step - loss: 1.8345 - val\_loss: 1.7005

Epoch 4/20

10/10 [=====] - ETA: 0s - loss: 1.6164f1\_score: 0.495 val\_f1\_s

core: 0.47666666666666667

10/10 [=====] - 1s 78ms/step - loss: 1.6164 - val\_loss: 1.5456

Epoch 5/20

10/10 [=====] - ETA: 0s - loss: 1.4814f1\_score: 0.5392857142857

143 val\_f1\_score: 0.5166666666666667

10/10 [=====] - 1s 80ms/step - loss: 1.4814 - val\_loss: 1.3876

Epoch 6/20

10/10 [=====] - ETA: 0s - loss: 1.3298f1\_score: 0.5285714285714

286 val\_f1\_score: 0.49166666666666664

10/10 [=====] - 1s 79ms/step - loss: 1.3298 - val\_loss: 1.3570

Epoch 7/20

10/10 [=====] - ETA: 0s - loss: 1.2606f1\_score: 0.605 val\_f1\_s

core: 0.61333333333333333

10/10 [=====] - 1s 78ms/step - loss: 1.2606 - val\_loss: 1.1905

Epoch 8/20

10/10 [=====] - ETA: 0s - loss: 1.1565f1\_score: 0.6321428571428

571 val\_f1\_score: 0.6416666666666667

10/10 [=====] - 1s 78ms/step - loss: 1.1565 - val\_loss: 1.1261

Epoch 9/20

10/10 [=====] - ETA: 0s - loss: 1.1304f1\_score: 0.6578571428571

428 val\_f1\_score: 0.64833333333333333

10/10 [=====] - 1s 79ms/step - loss: 1.1304 - val\_loss: 1.0641

Epoch 10/20

10/10 [=====] - ETA: 0s - loss: 1.0756f1\_score: 0.6242857142857

143 val\_f1\_score: 0.635

10/10 [=====] - 1s 80ms/step - loss: 1.0756 - val\_loss: 1.0906

Epoch 11/20

10/10 [=====] - ETA: 0s - loss: 0.9950f1\_score: 0.6957142857142

857 val\_f1\_score: 0.6916666666666667

10/10 [=====] - 1s 80ms/step - loss: 0.9950 - val\_loss: 0.9406

Epoch 12/20

10/10 [=====] - ETA: 0s - loss: 0.9270f1\_score: 0.7057142857142

857 val\_f1\_score: 0.7166666666666667

```

10/10 [=====] - 1s 79ms/step - loss: 0.9270 - val_loss: 0.9101
Epoch 13/20
10/10 [=====] - ETA: 0s - loss: 0.8696f1_score: 0.7421428571428
571 val_f1_score: 0.7433333333333333
10/10 [=====] - 1s 80ms/step - loss: 0.8696 - val_loss: 0.8404
Epoch 14/20
10/10 [=====] - ETA: 0s - loss: 0.8156f1_score: 0.7528571428571
43 val_f1_score: 0.7483333333333333
10/10 [=====] - 1s 79ms/step - loss: 0.8156 - val_loss: 0.8197
Epoch 15/20
10/10 [=====] - ETA: 0s - loss: 0.8364f1_score: 0.7485714285714
286 val_f1_score: 0.7383333333333333
10/10 [=====] - 1s 80ms/step - loss: 0.8364 - val_loss: 0.7908
Epoch 16/20
10/10 [=====] - ETA: 0s - loss: 0.7440f1_score: 0.7857142857142
857 val_f1_score: 0.7816666666666666
10/10 [=====] - 1s 79ms/step - loss: 0.7440 - val_loss: 0.7339
Epoch 17/20
10/10 [=====] - ETA: 0s - loss: 0.6876f1_score: 0.7842857142857
141 val_f1_score: 0.7583333333333333
10/10 [=====] - 1s 79ms/step - loss: 0.6876 - val_loss: 0.7013
Epoch 18/20
10/10 [=====] - ETA: 0s - loss: 0.6640f1_score: 0.8028571428571
428 val_f1_score: 0.7866666666666666
10/10 [=====] - 1s 79ms/step - loss: 0.6640 - val_loss: 0.6810
Epoch 19/20
10/10 [=====] - ETA: 0s - loss: 0.6423f1_score: 0.8035714285714
286 val_f1_score: 0.79
10/10 [=====] - 1s 92ms/step - loss: 0.6423 - val_loss: 0.6452
Epoch 20/20
10/10 [=====] - ETA: 0s - loss: 0.6234f1_score: 0.8207142857142
857 val_f1_score: 0.8000000000000000
10/10 [=====] - 1s 79ms/step - loss: 0.6234 - val_loss: 0.6283
<keras.callbacks.History at 0x7faa49df5e90>

```

Out[ ]:

```
In [ ]: %tensorboard --logdir logs/fits
```

```

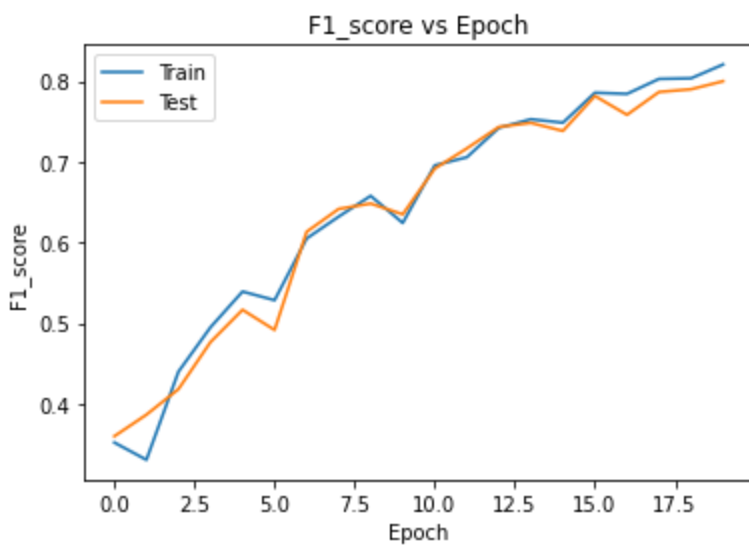
Reusing TensorBoard on port 6006 (pid 2003), started 0:20:01 ago. (Use '!kill 2003' to k
ill it.)

```

```

In [ ]: plt.plot(f1_fn.history['train_f1_score'])
plt.plot(f1_fn.history['val_f1_score'])
plt.title("F1_score vs Epoch")
plt.xlabel('Epoch')
plt.ylabel('F1_score')
plt.legend(["Train", "Test"])
plt.show()

```



### 3. Data augmentation with raw features

Till now we have done with 2000 samples only. It is very less data. We are giving the process of generating augmented data below.

There are two types of augmentation:

1. time stretching - Time stretching either increases or decreases the length of the file. For time stretching we move the file 30% faster or slower
2. pitch shifting - pitch shifting moves the frequencies higher or lower. For pitch shifting we shift up or down one half-step.

```
In [ ]: ## generating augmented data.
def generate_augmented_data(file_path):
    augmented_data = []
    samples = load_wav(file_path, get_duration=False)
    for time_value in [0.7, 1, 1.3]:
        for pitch_value in [-1, 0, 1]:
            time_stretch_data = librosa.effects.time_stretch(samples, rate=time_value)
            final_data = librosa.effects.pitch_shift(time_stretch_data, sr=sample_rate,
            augmented_data.append(final_data)
    return augmented_data
```

```
In [ ]: temp_path = df_audio.iloc[0].path
aug_temp = generate_augmented_data(temp_path)
```

```
In [ ]: len(aug_temp)
```

```
Out[ ]: 9
```

### Follow the steps

1. Split data 'df\_audio' into train and test (80-20 split)
2. We have 2000 data points(1600 train points, 400 test points)

```
In [ ]: X_train, X_test, y_train, y_test=train_test_split(df_audio['path'],df_audio['label'],ran
```



1. Do augmentation only on X\_train, pass each point of X\_train to generate\_augmented\_data function. After augmentation we will get 14400 train points. Make sure that you are augmenting the corresponding class labels (y\_train) also.
2. Preprocess your X\_test using load\_wav function.
3. Convert the augmented\_train\_data and test\_data to numpy arrays.
4. Perform padding and masking on augmented\_train\_data and test\_data.
5. After padding define the model similar to model 1 and fit the data

**Note** - While fitting your model on the augmented data for model 3 you might face Resource exhaust error. One simple hack to avoid that is save the augmented\_train\_data, augment\_y\_train, test\_data and y\_test to Drive or into your local system. Then restart the runtime so that now you can train your model with full RAM capacity. Upload these files again in the new runtime session perform padding and masking and then fit your model.

```
In [ ]: train_augmented_data = []
        train_labels = []

        for path, label in zip(X_train.values, y_train.values):
            augmented_data = generate_augmented_data(path)
            train_augmented_data.extend(augmented_data)
            train_labels.extend(label*9)

X_train_processed = pd.DataFrame({'raw_data' : train_augmented_data, 'label' : train_labels})

test_data = []
for path in X_test.values:
    test_data.append(load_wav(path, get_duration = False))

X_test_processed = pd.DataFrame({'raw_data' : test_data, 'label' : y_test.values})
```

```
In [ ]: #padding the sequences
X_train_pad_seq = pad_sequences(X_train_processed['raw_data'], maxlen = max_length,
                                dtype = 'float64', padding = 'post',
                                truncating = 'post')
X_test_pad_seq = pad_sequences(X_test_processed['raw_data'], maxlen = max_length,
                                dtype = 'float64', padding = 'post',
                                truncating = 'post')

y_train = X_train_processed.label.values.astype('int')
y_test = X_test_processed.label.values.astype('int')

X_train_mask = np.array(X_train_pad_seq != 0)
X_test_mask = np.array(X_test_pad_seq != 0)

X_train_pad_seq = np.expand_dims(X_train_pad_seq, -1)
X_test_pad_seq = np.expand_dims(X_test_pad_seq, -1)
```

```
In [ ]: input_layer = Input(shape = (17640,1))
input_mask = Input(shape = (17640), dtype = 'bool')
lstm_layer = LSTM(50)(input_layer, mask = input_mask)
dense_layer = Dense(100, activation = 'relu', kernel_initializer = tf.keras.initializers
output_layer = Dense(10, activation = 'softmax')(dense_layer)

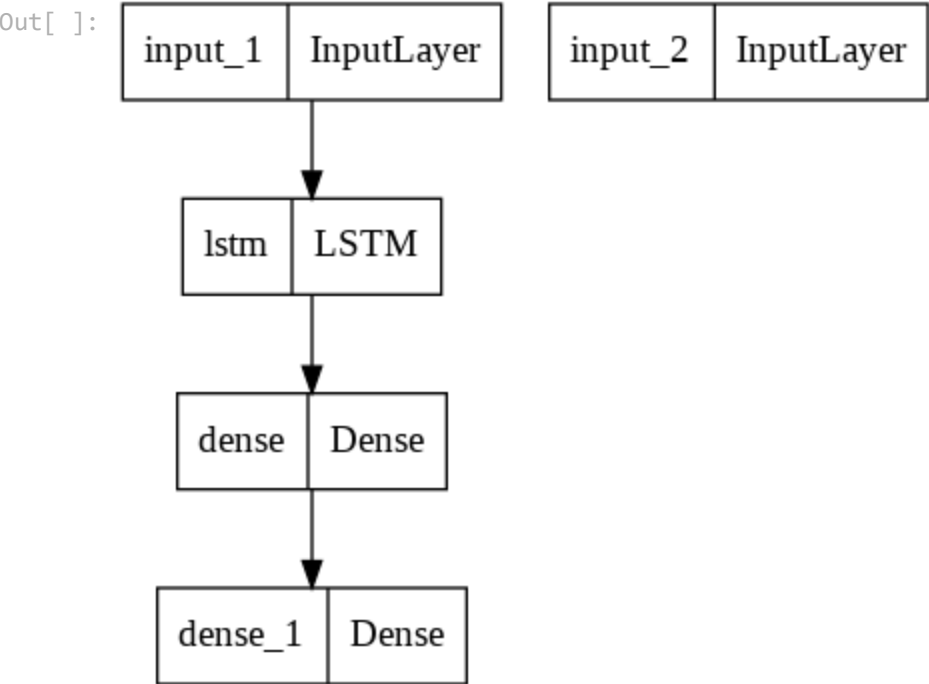
model3 = Model(inputs = [input_layer, input_mask], outputs = output_layer)

model3.summary()
```

Model: "model"

Layer (type)	Output Shape	Param #	Connected to
=====			
input_1 (InputLayer)	[(None, 17640, 1)]	0	[]
input_2 (InputLayer)	[(None, 17640)]	0	[]
lstm (LSTM)	(None, 50)	10400	['input_1[0][0]', 'input_2[0][0]']
dense (Dense)	(None, 100)	5100	['lstm[0][0]']
dense_1 (Dense)	(None, 10)	1010	['dense[0][0]']
=====			
Total params: 16,510			
Trainable params: 16,510			
Non-trainable params: 0			

```
In [ ]: tf.keras.utils.plot_model(
    model3, to_file='model_3.png', show_shapes=False, show_layer_names=True,
    rankdir='TB', expand_nested=False, dpi=96
)
```



```
In [ ]: model3.compile(optimizer = tf.keras.optimizers.Adam(0.006), loss = 'sparse_categorical_crossentropy')
```

```
In [ ]: tf.keras.backend.clear_session()
```

```
!rm -rf ./logs/
```

```
In [ ]: %load_ext tensorboard
import datetime, os
from tensorflow.keras.callbacks import ModelCheckpoint

f1_fn = Metric_f1([X_train_pad_seq,X_train_mask], y_train, ([X_test_pad_seq, X_test_mask

log_dir = os.path.join("logs",'fits', datetime.datetime.now().strftime("%Y%m%d-%H%M%S"))
tensorboard_callback = tf.keras.callbacks.TensorBoard(log_dir = log_dir,histogram_freq=1
model3.fit(x = [X_train_pad_seq,X_train_mask], y = y_train, epochs = 5, verbose = 1, bat
validation_data = ([X_test_pad_seq, X_test_mask], y_test) ,
callbacks = [tensorboard_callback, f1_fn])
```

The tensorboard extension is already loaded. To reload it, use:

```
%reload_ext tensorboard
```

Epoch 1/5

```
144/144 [=====] - ETA: 0s - loss: 2.3039f1_score: 0.10000000000
000002 val_f1_score: 0.10000000000000002
144/144 [=====] - 211s 1s/step - loss: 2.3039 - val_loss: 2.302
9
```

Epoch 2/5

```
144/144 [=====] - ETA: 0s - loss: 2.3033f1_score: 0.10000000000
000002 val_f1_score: 0.10000000000000002
144/144 [=====] - 204s 1s/step - loss: 2.3033 - val_loss: 2.302
7
```

Epoch 3/5

```
144/144 [=====] - ETA: 0s - loss: 2.3033f1_score: 0.10000000000
000002 val_f1_score: 0.10000000000000002
144/144 [=====] - 203s 1s/step - loss: 2.3033 - val_loss: 2.302
9
```

Epoch 4/5

```
144/144 [=====] - ETA: 0s - loss: 2.3033f1_score: 0.10000000000
000002 val_f1_score: 0.10000000000000002
144/144 [=====] - 204s 1s/step - loss: 2.3033 - val_loss: 2.302
9
```

Epoch 5/5

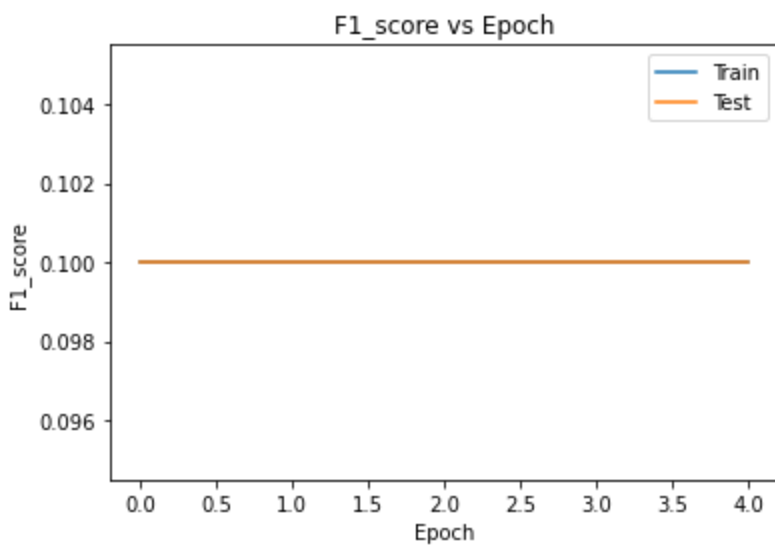
```
144/144 [=====] - ETA: 0s - loss: 2.3032f1_score: 0.10000000000
000002 val_f1_score: 0.10000000000000002
144/144 [=====] - 204s 1s/step - loss: 2.3032 - val_loss: 2.302
8
```

```
Out[ ]: <keras.callbacks.History at 0x7faa5fb920d0>
```

```
In [ ]: %tensorboard --logdir logs/fits
```

```
Reusing TensorBoard on port 6006 (pid 2003), started 0:57:48 ago. (Use '!kill 2003' to k
ill it.)
```

```
In [ ]: plt.plot(f1_fn.history['train_f1_score'])
plt.plot(f1_fn.history['val_f1_score'])
plt.title("F1_score vs Epoch")
plt.xlabel('Epoch')
plt.ylabel('F1_score')
plt.legend(["Train", "Test"])
plt.show()
```



## 4. Data augmentation with spectrogram data

1. use `convert_to_spectrogram` and convert the padded data from train and test data to spectrogram data.
2. The shape of train data will be 14400 x 64 x 35 and shape of test\_data will be 400 x 64 x 35
3. Define the model similar to model 2 and fit the data

```
In [ ]: X_train_pad_seq = np.squeeze(X_train_pad_seq, axis = -1)
X_test_pad_seq = np.squeeze(X_test_pad_seq, axis = -1)
```

```
In [ ]: X_train_spectrogram = []
for x in X_train_pad_seq:
    X_train_spectrogram.append(convert_to_spectrogram(x))

X_train_spectrogram = np.array(X_train_spectrogram)

X_test_spectrogram = []
for x in X_test_pad_seq:
    X_test_spectrogram.append(convert_to_spectrogram(x))

X_test_spectrogram = np.array(X_test_spectrogram)
```

```
In [ ]: input_layer = Input(shape = (64,35,))
lstm_layer = LSTM(512, return_sequences = True)(input_layer)
x = tf.math.reduce_mean(lstm_layer, axis = -1)
dense_layer = Dense(512, activation = 'relu', kernel_initializer = tf.keras.initializers
output_layer = Dense(10, activation = 'softmax')(dense_layer)

model4 = Model(inputs = input_layer, outputs = output_layer)
#printing the model summary
model4.summary()
```

Model: "model"

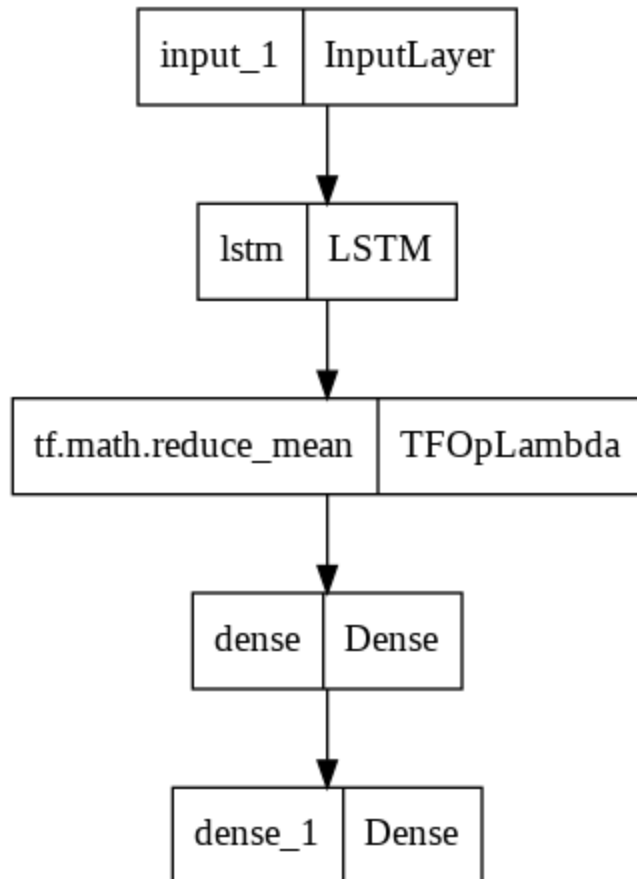
Layer (type)	Output Shape	Param #
=====		
input_1 (InputLayer)	[(None, 64, 35)]	0
lstm (LSTM)	(None, 64, 512)	1122304
tf.math.reduce_mean (TFOpLa mbda)	(None, 64)	0
dense (Dense)	(None, 512)	33280

dense\_1 (Dense) (None, 10) 5130

```
=====
Total params: 1,160,714
Trainable params: 1,160,714
Non-trainable params: 0
=====
```

```
In [ ]: tf.keras.utils.plot_model(
        model4, to_file='model_4.png', show_shapes=False, show_layer_names=True,
        rankdir='TB', expand_nested=False, dpi=96
    )
```

Out[ ]:



```
In [ ]: model4.compile(optimizer = tf.keras.optimizers.Adam(0.001), loss = 'sparse_categorical_crossentropy')
```

```
In [ ]: tf.keras.backend.clear_session()
!rm -rf ./logs/
```

```
In [ ]: %load_ext tensorboard
import datetime, os
from tensorflow.keras.callbacks import ModelCheckpoint

f1_fn = Metric_f1([X_train_spectrogram], y_train, ([X_test_spectrogram], y_test))

log_dir = os.path.join("logs", 'fits', datetime.datetime.now().strftime("%Y%m%d-%H%M%S"))
tensorboard_callback = tf.keras.callbacks.TensorBoard(log_dir = log_dir, histogram_freq=1)
model4.fit(x = [X_train_spectrogram], y = y_train, epochs = 20, verbose = 1, batch_size = 128,
          validation_data = ([X_test_spectrogram], y_test) ,
          callbacks = [tensorboard_callback, f1_fn])
```

The tensorboard extension is already loaded. To reload it, use:

```
%reload_ext tensorboard
```

Epoch 1/20

4/103 [>.....] - ETA: 2s - loss: 2.3010

WARNING:tensorflow:Callback method `on\_train\_batch\_end` is slow compared to the batch training step. Please see the TensorFlow documentation for details. Called from: C:\Program Files\Python36\lib\site-packages\tensorflow\keras\tensorboard\tensorboard.py:101

```
me (batch time: 0.0101s vs `on_train_batch_end` time: 0.0160s). Check your callbacks.
103/103 [=====] - ETA: 0s - loss: 1.4955f1_score: 0.72354166666
66666 val_f1_score: 0.8075
103/103 [=====] - 7s 50ms/step - loss: 1.4955 - val_loss: 0.748
8
Epoch 2/20
103/103 [=====] - ETA: 0s - loss: 0.6760f1_score: 0.82930555555
55554 val_f1_score: 0.8725
103/103 [=====] - 4s 40ms/step - loss: 0.6760 - val_loss: 0.440
4
Epoch 3/20
103/103 [=====] - ETA: 0s - loss: 0.4723f1_score: 0.86520833333
33334 val_f1_score: 0.915
103/103 [=====] - 4s 40ms/step - loss: 0.4723 - val_loss: 0.334
5
Epoch 4/20
103/103 [=====] - ETA: 0s - loss: 0.4099f1_score: 0.88256944444
44445 val_f1_score: 0.9075
103/103 [=====] - 4s 40ms/step - loss: 0.4099 - val_loss: 0.302
0
Epoch 5/20
103/103 [=====] - ETA: 0s - loss: 0.3717f1_score: 0.89215277777
77778 val_f1_score: 0.915
103/103 [=====] - 4s 40ms/step - loss: 0.3717 - val_loss: 0.281
4
Epoch 6/20
103/103 [=====] - ETA: 0s - loss: 0.3212f1_score: 0.90833333333
33333 val_f1_score: 0.9325
103/103 [=====] - 4s 40ms/step - loss: 0.3212 - val_loss: 0.243
3
Epoch 7/20
103/103 [=====] - ETA: 0s - loss: 0.2961f1_score: 0.91423611111
11111 val_f1_score: 0.9275
103/103 [=====] - 4s 40ms/step - loss: 0.2961 - val_loss: 0.236
2
Epoch 8/20
103/103 [=====] - ETA: 0s - loss: 0.2850f1_score: 0.92055555555
55556 val_f1_score: 0.9325
103/103 [=====] - 4s 40ms/step - loss: 0.2850 - val_loss: 0.215
0
Epoch 9/20
103/103 [=====] - ETA: 0s - loss: 0.2621f1_score: 0.92680555555
55557 val_f1_score: 0.935
103/103 [=====] - 4s 39ms/step - loss: 0.2621 - val_loss: 0.205
7
Epoch 10/20
103/103 [=====] - ETA: 0s - loss: 0.2442f1_score: 0.93458333333
33333 val_f1_score: 0.945
103/103 [=====] - 4s 40ms/step - loss: 0.2442 - val_loss: 0.191
7
Epoch 11/20
103/103 [=====] - ETA: 0s - loss: 0.2324f1_score: 0.93520833333
33332 val_f1_score: 0.9375
103/103 [=====] - 4s 41ms/step - loss: 0.2324 - val_loss: 0.180
5
Epoch 12/20
103/103 [=====] - ETA: 0s - loss: 0.2180f1_score: 0.93548611111
11111 val_f1_score: 0.94
103/103 [=====] - 4s 41ms/step - loss: 0.2180 - val_loss: 0.176
2
Epoch 13/20
103/103 [=====] - ETA: 0s - loss: 0.2093f1_score: 0.93326388888
88889 val_f1_score: 0.9425
103/103 [=====] - 4s 39ms/step - loss: 0.2093 - val_loss: 0.170
4
```

```

Epoch 14/20
103/103 [=====] - ETA: 0s - loss: 0.1971f1_score: 0.9461111111
11111 val_f1_score: 0.9525
103/103 [=====] - 4s 39ms/step - loss: 0.1971 - val_loss: 0.171
6
Epoch 15/20
103/103 [=====] - ETA: 0s - loss: 0.1918f1_score: 0.9445833333
33333 val_f1_score: 0.955
103/103 [=====] - 4s 40ms/step - loss: 0.1918 - val_loss: 0.155
8
Epoch 16/20
103/103 [=====] - ETA: 0s - loss: 0.1776f1_score: 0.9477083333
33334 val_f1_score: 0.9500000000000001
103/103 [=====] - 4s 40ms/step - loss: 0.1776 - val_loss: 0.164
0
Epoch 17/20
103/103 [=====] - ETA: 0s - loss: 0.1719f1_score: 0.9529166666
66666 val_f1_score: 0.9525
103/103 [=====] - 4s 40ms/step - loss: 0.1719 - val_loss: 0.143
9
Epoch 18/20
103/103 [=====] - ETA: 0s - loss: 0.1721f1_score: 0.9536805555
55555 val_f1_score: 0.9675
103/103 [=====] - 4s 40ms/step - loss: 0.1721 - val_loss: 0.139
0
Epoch 19/20
103/103 [=====] - ETA: 0s - loss: 0.1606f1_score: 0.9470833333
33334 val_f1_score: 0.955
103/103 [=====] - 4s 40ms/step - loss: 0.1606 - val_loss: 0.167
7
Epoch 20/20
101/103 [=====>.] - ETA: 0s - loss: 0.1660f1_score: 0.9458333333
33333 val_f1_score: 0.945
103/103 [=====] - 4s 41ms/step - loss: 0.1665 - val_loss: 0.156
7

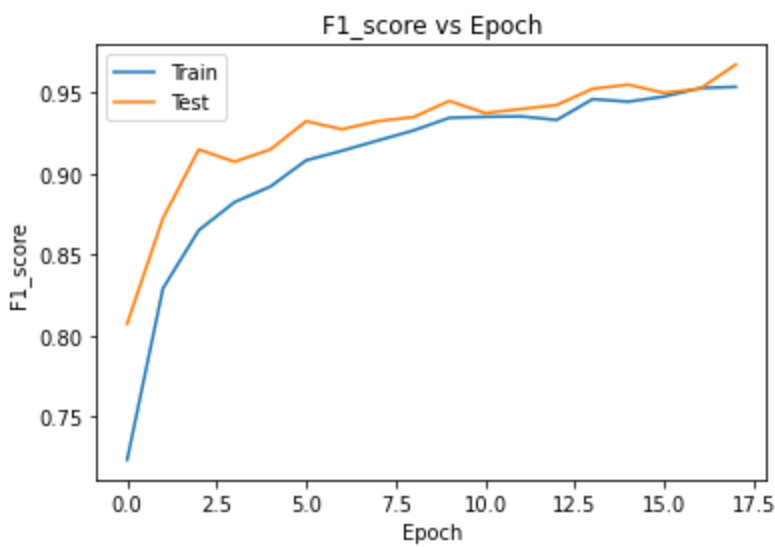
```

```
Out[ ]: <keras.callbacks.History at 0x7faa49229150>
```

```
In [ ]: %tensorboard --logdir logs/fits
```

```
Reusing TensorBoard on port 6006 (pid 2003), started 1:05:09 ago. (Use '!kill 2003' to k
ill it.)
```

```
In [ ]: plt.plot(f1_fn.history['train_f1_score'][:18])
plt.plot(f1_fn.history['val_f1_score'][:18])
plt.title("F1_score vs Epoch")
plt.xlabel('Epoch')
plt.ylabel('F1_score')
plt.legend(["Train", "Test"])
plt.show()
```



Observations:

1. The use of spectrogram of audio data for modelling is preferred over raw audio as with spectrogram data we can achieve very good f1 score in low no. of epochs. Even using augmentation to increase the dataset size doesn't help the performance of models using raw data inputs.
2. For augmentation we shifted the data by 30% on the time axis on both directions and even decreased and increased the pitch of the audio.
3. Augmentation of data helped improve the performance of model using spectrogram data.

```
In [170... from prettytable import PrettyTable
x = PrettyTable()

x.field_names = ["Model", "Augemented", "Train F1 Score", "validation F1 Score"]
x.add_row(["model 1", "No", 0.1016, 0.1007])
x.add_row(["model 2", "Yes", 0.8207, 0.8])
x.add_row(["model 3", "No", 0.1, 0.1])
x.add_row(["model 4", "Yes", 0.9537, 0.9675])
print(x)
```

Model	Augemented	Train F1 Score	validation F1 Score
model 1	No	0.1016	0.1007
model 2	Yes	0.8207	0.8
model 3	No	0.1	0.1
model 4	Yes	0.9537	0.9675