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 instrctions are given below. You have to write the code in the same cell which contains the instrction.
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.

Dataset Link

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.

Libraries and versions

```
pip install -U pip
pip install -U pandas==1.5.0
pip install -U numpy==1.23.4
pip install -U seaborn==0.12.0
pip install -U tqdm==4.64.1
pip install -U tensorflow==2.10.0
pip install -U scikit-learn==1.1.2
pip install -U librosa==0.9.2
pip install -U pydot==1.4.2
pip install -U prettytable==3.4.1
pip install -U plotly-express==0.4.1
```

In [1]:

```
import os
import pickle
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.graph_objects as go
from plotly.subplots import make_subplots

from tqdm import tqdm
from datetime import datetime
from prettytable import FRAME
from prettytable import PrettyTable
from prettytable import SINGLE_BORDER
from IPython.display import Audio
from IPython.display import YouTubeVideo

from sklearn.utils import shuffle
from sklearn.metrics import f1_score
from sklearn.model_selection import train_test_split

import librosa
import librosa.display

import tensorflow as tf
```

```python
from tensorflow.keras.layers import LSTM
from tensorflow.keras.models import Model
from tensorflow.keras.layers import Input
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Dropout
from tensorflow.keras.layers import BatchNormalization
from tensorflow.keras.layers import GlobalAveragePooling1D
from tensorflow.keras.utils import pad_sequences
from tensorflow.keras.callbacks import Callback
from tensorflow.keras.callbacks import TensorBoard
from tensorflow.keras.callbacks import ModelCheckpoint
from tensorflow.keras.callbacks import ReduceLROnPlateau

import warnings
warnings.filterwarnings('ignore')

plt.style.use('fivethirtyeight')
```

```
2022-11-08 23:40:38.193490: I tensorflow/core/platform/cpu_feature_guard.cc:193] This TensorFlow binary is optimi
zed with oneAPI Deep Neural Network Library (oneDNN) to use the following CPU instructions in performance-critica
l operations:  AVX2 FMA
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
2022-11-08 23:40:38.327186: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynami
c library 'libcudart.so.11.0'; dlerror: libcudart.so.11.0: cannot open shared object file: No such file or direct
ory
2022-11-08 23:40:38.327213: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cudart dlerror if y
ou do not have a GPU set up on your machine.
2022-11-08 23:40:38.355441: E tensorflow/stream_executor/cuda/cuda_blas.cc:2981] Unable to register cuBLAS factor
y: Attempting to register factory for plugin cuBLAS when one has already been registered
2022-11-08 23:40:38.922750: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynami
c library 'libnvinfer.so.7'; dlerror: libnvinfer.so.7: cannot open shared object file: No such file or directory
2022-11-08 23:40:38.922810: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynami
c library 'libnvinfer_plugin.so.7'; dlerror: libnvinfer_plugin.so.7: cannot open shared object file: No such file
or directory
2022-11-08 23:40:38.922815: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Cannot dlopen
some TensorRT libraries. If you would like to use Nvidia GPU with TensorRT, please make sure the missing librarie
s mentioned above are installed properly.
```

We shared recordings.zip, please unzip those.

In [2]:

```python
# 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"

DIR_NAME = 'recordings/'
all_files = [DIR_NAME + file for file in os.listdir(DIR_NAME)]
print(f'Number of files recorded : {len(all_files)}')
```

```
Number of files recorded : 2000
```

Grader function 1

In [3]:

```python
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[3]:

```
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

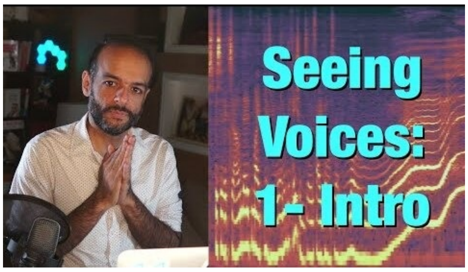
Exploring the sound dataset

In [4]:

```python
# It is a good programming practise to explore the dataset that you are dealing with.
# This dataset is unique in itself because it has sounds as input
# https://colab.research.google.com/github/Tyler-Hilbert/AudioProcessingInPythonWorkshop/blob/master/AudioProces
```

```python
# visualize the data and write code to play 2-3 sound samples in the notebook for better understanding.
# please go through the following reference video https://www.youtube.com/watch?v=37zCgCdV468

YT_videos = ['https://youtu.be/37zCgCdV468', 'https://youtu.be/m3XbqfIij_Y',
             'https://youtu.be/0a_d-zaUti8', 'https://youtu.be/ZqpSb5p1xQo']

print('\nLink to the reference videos listed bellow\n', '-' * 41)
for idx, video in enumerate(YT_videos):
    print(f'{idx+1}. {video}')
```

```
Link to the reference videos listed bellow
 -----------------------------------------
1. https://youtu.be/37zCgCdV468
2. https://youtu.be/m3XbqfIij_Y
3. https://youtu.be/0a_d-zaUti8
4. https://youtu.be/ZqpSb5p1xQo
```

1. 

2. 

3. 

4. 

In [5]:

```python
# Lets listen few random audio files from the collection
# https://plotly.com/python/subplots/

random_list = np.random.choice(all_files, size = 5, replace = False)

fig = go.Figure()
fig = make_subplots(rows = 5)
for idx, audio in enumerate(random_list):

    print(audio.split('/')[-1])
    display(Audio(audio))
    raw_data, sr = librosa.load(audio)
    fig.add_scatter(y = raw_data, row = idx+1, col = 1, name = audio.split('/')[-1])

fig.update_layout(height = 900, width = 850, title_text = 'Wave form of sampled audio files')
fig.show()
```

1_nicolas_32.wav

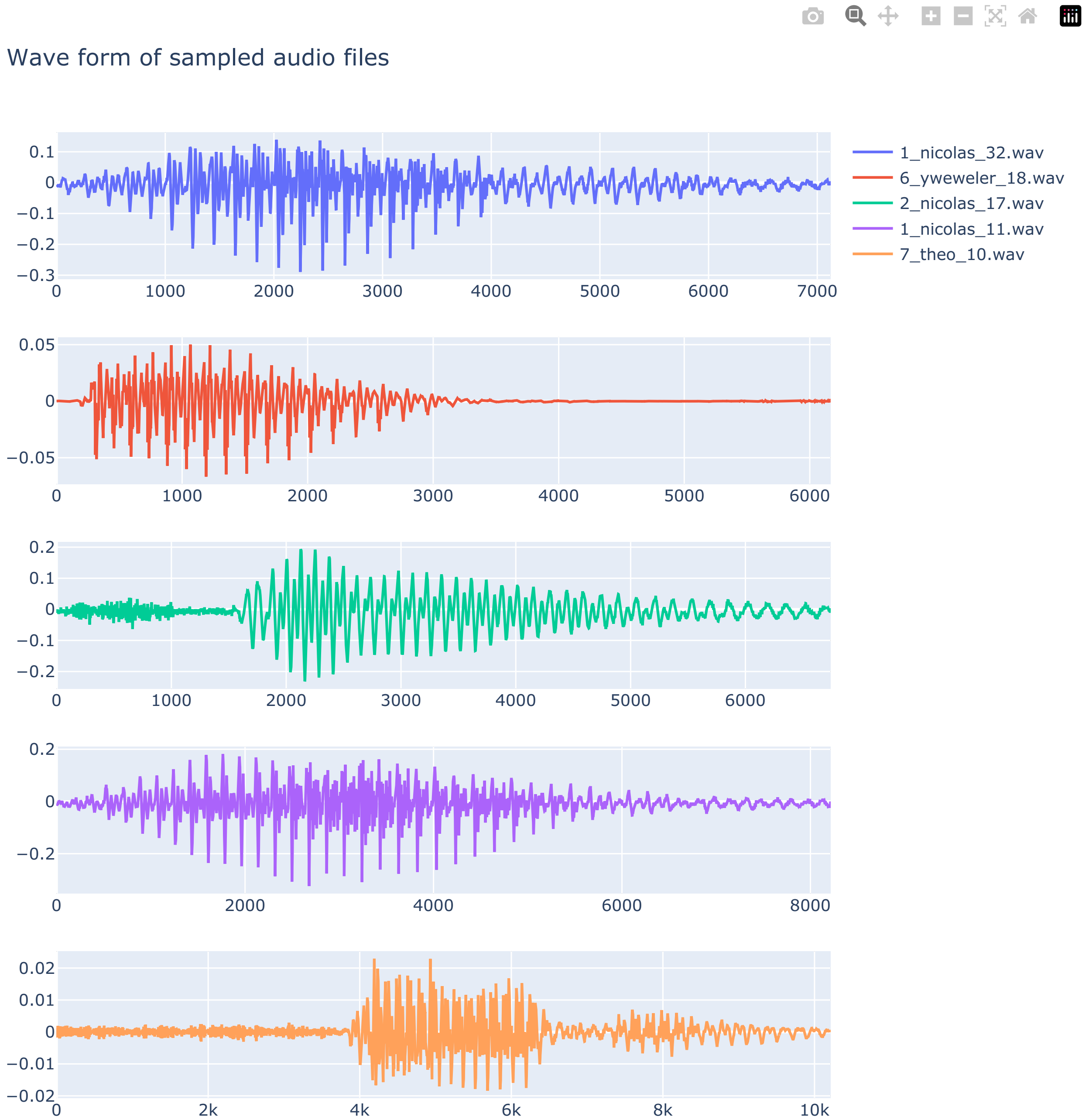
6_yweweler_18.wav

2_nicolas_17.wav

1_nicolas_11.wav

7_theo_10.wav

7_theo_10.wav



In [6]:

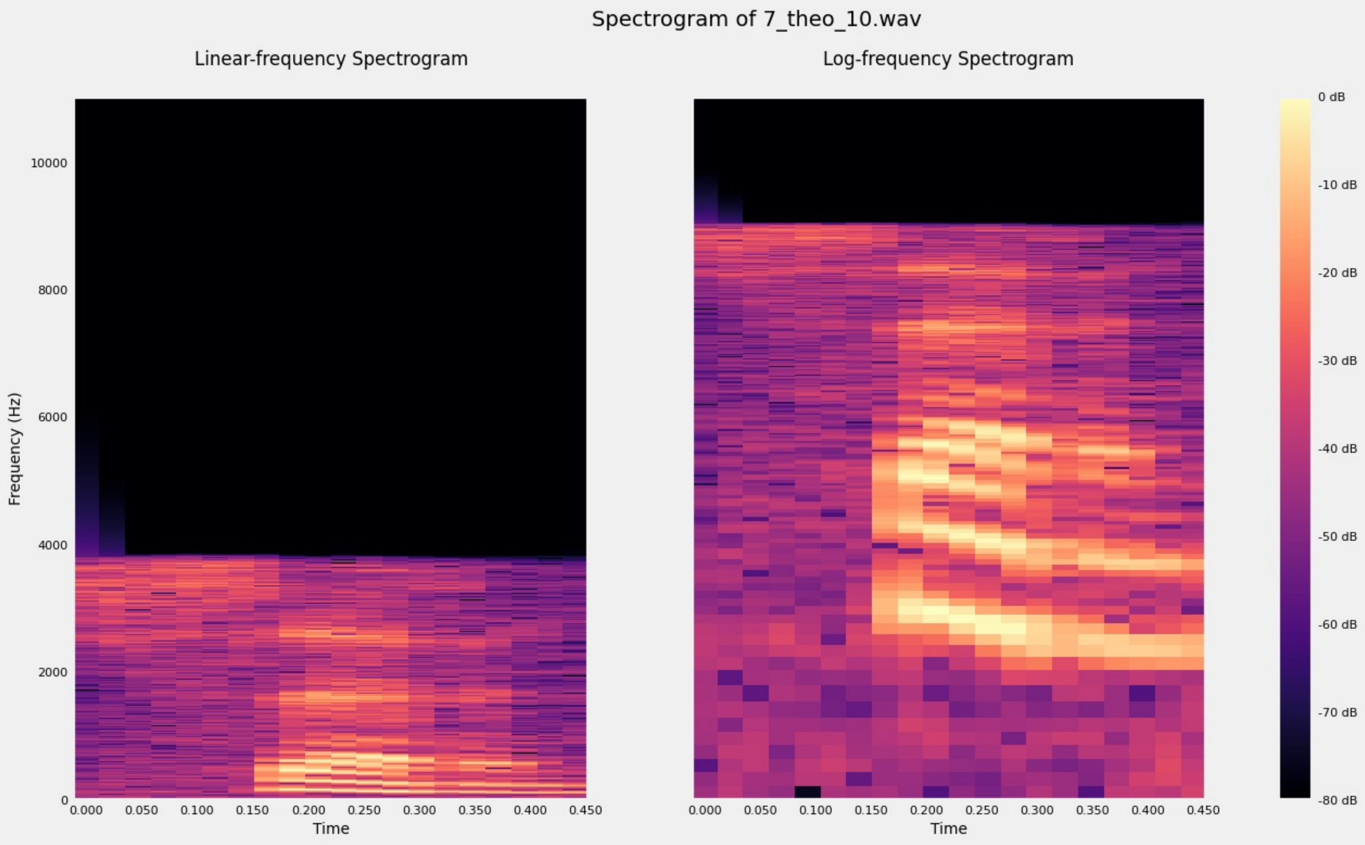
```
# Convet wave form to Frequency-Time format

D = librosa.stft(raw_data) # Short-time Fourier transform (STFT)
S_db = librosa.amplitude_to_db(np.abs(D), ref = np.max)
print(f"Shape of 'S_db' :: {S_db.shape}")
```

```
Shape of 'S_db' :: (1025, 20)
```

In [7]:

```
# https://stackoverflow.com/a/39566040
# https://librosa.org/doc/main/generated/librosa.display.specshow.html

plt.rc('xtick', labelsize = 8)
plt.rc('ytick', labelsize = 8)
plt.rc('axes', labelsize = 10)
plt.rc('axes', titlesize = 12)

fig, ax = plt.subplots(nrows = 1, ncols = 2, sharex = True, figsize = (15, 8))
fig.suptitle(f"Spectrogram of {audio.split('/')[-1]}\n", fontsize = 14)
```

```
img = librosa.display.specshow(S_db, y_axis = 'linear', x_axis = 'time', ax = ax[0])
ax[0].set_title('Linear-frequency Spectrogram\n')
ax[0].set_ylabel('Frequency (Hz)')

img = librosa.display.specshow(S_db, y_axis = 'log', x_axis = 'time', ax = ax[1])
ax[1].set_title('Log-frequency Spectrogram\n')
ax[1].set_ylabel('Frequency (Hz)')
ax[1].label_outer()

fig.colorbar(img, ax = ax, format = '%.f dB')

plt.show()
```



Creating dataframe

In [8]:

```
# 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

labels = [file.split('/')[-1][-0] for file in all_files]
df_audio = pd.DataFrame({'path' : all_files, 'label' : labels})

df_audio.head()
```

Out[8]:

	path	label
0	recordings/0_jackson_44.wav	0
1	recordings/9_jackson_36.wav	9
2	recordings/8_nicolas_41.wav	8
3	recordings/3_jackson_19.wav	3
4	recordings/3_yweweler_21.wav	3

In [9]:

```
# 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 [10]:
```
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[10]:
```
True
```

In [11]:
```
df_audio = shuffle(df_audio, random_state = 33) # don't change the random state
df_audio.head()
```

Out[11]:

	path	label
766	recordings/0_nicolas_39.wav	0
182	recordings/5_nicolas_15.wav	5
1763	recordings/5_jackson_6.wav	5
1814	recordings/7_nicolas_33.wav	7
596	recordings/6_nicolas_35.wav	6

Train and Validation split

In [12]:
```
# 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%

X = df_audio.path
Y = df_audio.label

X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size = 0.3,
                                                    stratify = Y, random_state = 45)
```

Grader function 3

In [13]:
```
def grader_split():

    flag_len = (len(X_train) == 1400) and (len(X_test) == 600) and \
                                    (len(y_train) == 1400) and (len(y_test) == 600)

    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[13]:
```
True
```

Preprocessing

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

In [14]:
```python
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:

        # From version 0.10 passing these as positional arguments will result in an error
        # librosa.get_duration(samples, sample_rate)
        duration = librosa.get_duration(y = samples, sr = sample_rate)
        return [samples, duration]

    else:
        return samples
```

In [15]:
```python
# 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, duration),
# with same index of X_train/y_train

def process_wave_data(files, name):

    duration, sample = [], []

    for file in tqdm(files, name + ' '):
        sample_ , duration_ = load_wav(file)
        sample.append(sample_)
        duration.append(duration_)

    return pd.DataFrame({'raw_data' : sample, 'duration' : duration})

X_train_processed = process_wave_data(X_train, 'X_train')
X_test_processed = process_wave_data(X_test, 'X_test')
```

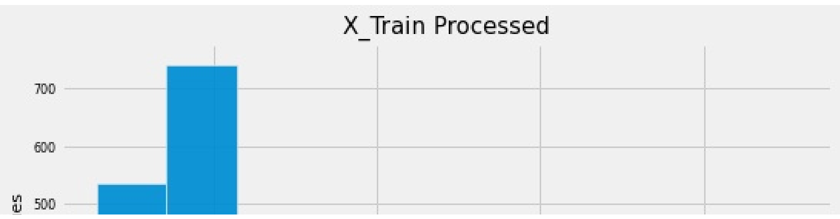
```
X_train : 100%|██████████| 1400/1400 [00:15<00:00, 91.75it/s]
X_test : 100%|██████████| 600/600 [00:05<00:00, 105.73it/s]
```

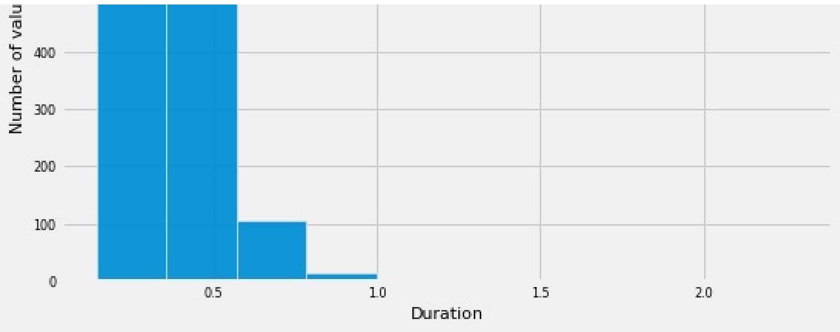
In [16]:
```python
X_train_processed.head()
```

Out[16]:

	raw_data	duration
0	[0.0058789197, 0.0017868795, -0.0027949458, -0...	0.796281
1	[0.00031183258, 0.00040326704, 0.0004383796, 0...	0.215011
2	[0.0032379585, 0.0011577337, -0.00018193743, 0...	0.678503
3	[0.00046396037, 0.00043442327, 0.0002750368, 6...	0.430884
4	[-0.010306078, -0.012605253, -0.013264463, -0....	0.597778

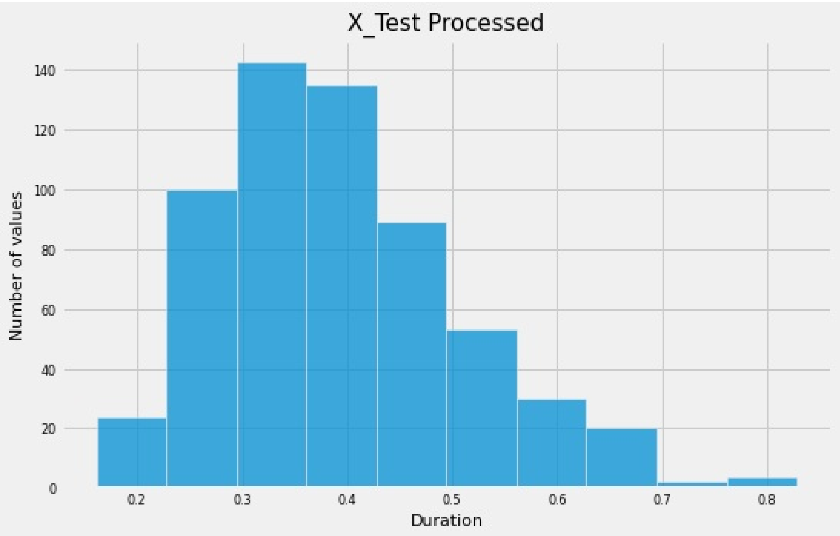
In [17]:
```python
# plot the histogram of the duration for trian

plt.rc('axes', titlesize = 15)
plt.rc('axes', labelsize = 11)

plt.rcParams['figure.figsize'] = [8, 5]

ax = sns.histplot(X_train_processed.duration, bins = 10)

ax.set_xlabel('Duration')
ax.set_ylabel('Number of values')
ax.set_title('X_Train Processed')
sns.histplot(X_train_processed.duration, bins = 10)
plt.show()
```





In [18]:

```
# plot the histogram of the duration for test

ax = sns.histplot(X_test_processed.duration,bins = 10)

ax.set_xlabel('Duration')
ax.set_ylabel('Number of values')
ax.set_title('X_Test Processed')

plt.show()
```



In [19]:

```
# Percentile print function

def percentiles(df, lower, upper, incerment):
     for val in range(lower, upper + incerment, incerment):
            print(f'{val:3d}th percentile is {np.percentile(df.duration, val)}')
```

In [20]:

```
# print 0 to 100 percentile values with step size of 10 for train data duration.

percentiles(X_train_processed, 0, 100, 10)
```

```
  0th percentile is 0.1435374149659864
 10th percentile is 0.263750566893424
 20th percentile is 0.30201360544217687
 30th percentile is 0.3361043083900227
 40th percentile is 0.36248526077097504
 50th percentile is 0.39378684807256237
 60th percentile is 0.42107936507936505
 70th percentile is 0.45006349206349205
 80th percentile is 0.4850702947845805
 90th percentile is 0.5581768707482994
100th percentile is 2.282766439909297
```

In [21]:

```
# print 90 to 100 percentile values with step size of 1.

percentiles(X_train_processed, 90, 100, 1)
```

```
 90th percentile is 0.5581768707482994
 91th percentile is 0.5704222222222223
 92th percentile is 0.581734240362812
 93th percentile is 0.6012321995464855
 94th percentile is 0.617911111111111
```

```
 95th percentile is 0.6330226757369615
 96th percentile is 0.6459265306122448
 97th percentile is 0.6743265306122448
 98th percentile is 0.7174285714285713
 99th percentile is 0.8209596371882086
100th percentile is 2.282766439909297
```

Grader function 4

In [22]:
```
def grader_processed():

    flag_columns = (all(X_train_processed.columns == ['raw_data', 'duration'])) and \
                                        (all(X_test_processed.columns == ['raw_data', 'duration']))

    flag_shape = (X_train_processed.shape == (1400, 2)) and (X_test_processed.shape == (600,2))

    return flag_columns and flag_shape

grader_processed()
```

Out[22]:
```
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*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 [23]:
```
max_length  = 17640
```

In [24]:
```
# as discussed above, Pad with Zero if length of sequence is less than 17640 else Truncate the number.
# 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 arrays mask vector,
# dtype must be bool.

# https://www.tensorflow.org/api_docs/python/tf/keras/utils/pad_sequences

X_train_pad_seq = pad_sequences(X_train_processed.raw_data, maxlen = max_length, padding = 'post',
                                                    value = 0, dtype = 'float', truncating = 'post')
X_test_pad_seq = pad_sequences(X_test_processed.raw_data, maxlen = max_length, padding = 'post',
                                                   value = 0, dtype = 'float', truncating = 'post')

print(f"Shape of 'X_train_pad_seq' :: {X_train_pad_seq.shape}")
print(f"Shape of 'X_test_pad_seq'  :: {X_test_pad_seq.shape}")

X_train_mask = X_train_pad_seq.astype('bool')
X_test_mask = X_test_pad_seq.astype('bool')
```

```
Shape of 'X_train_pad_seq' :: (1400, 17640)
Shape of 'X_test_pad_seq'  :: (600, 17640)
```

Grader function 5

In [25]:
```
def grader_padoutput():

    flag_padshape = (X_train_pad_seq.shape==(1400, 17640)) and (X_test_pad_seq.shape==(600, 17640)) and \
                    (y_train.shape==(1400,))

    flag_maskshape = (X_train_mask.shape==(1400, 17640)) and (X_test_mask.shape==(600, 17640)) and \
                     (y_test.shape==(600,))

    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[25]:
```
True
```

Acceptance Criteria

Model	Micro F1 score
Model 1 & Model 3	0.10
Model 2 & Model 4	0.80

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) in tensorflow to know more about mask and also 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 [26]:
```
if not os.path.isdir('results'):
    os.mkdir('results')
```

In [27]:
```
y_train_int = y_train.astype('int')
y_test_int = y_test.astype('int')
```

In [28]:
```
# 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

# https://keras.io/api/models/model/

def model_1_3(max_length, model_name):

    tf.keras.backend.clear_session()

    input_pad = Input(shape = (max_length, 1))
    input_mask = Input(shape = max_length, dtype = 'bool')
    x = LSTM(25)(input_pad, mask = input_mask)
    x = Dense(50, activation = 'relu')(x)
    x = Dropout(0.5)(x)
    x = BatchNormalization()(x)
    x = Dense(40, activation = 'relu')(x)
    x = Dropout(0.3)(x)
    output_ = Dense(10, activation = 'softmax')(x)

    return Model(inputs = [input_pad, input_mask], outputs = output_, name = model_name)
```

```
In [29]: model_raw = model_1_3(max_length, 'RAW_Data_Alone')

model_raw.summary()
```

```
Model: "RAW_Data_Alone"
__________________________________________________________________________________________________
 Layer (type)                   Output Shape         Param #     Connected to
==================================================================================================
 input_1 (InputLayer)           [(None, 17640, 1)]   0           []

 input_2 (InputLayer)           [(None, 17640)]      0           []

 lstm (LSTM)                    (None, 25)           2700        ['input_1[0][0]',
                                                                  'input_2[0][0]']

 dense (Dense)                  (None, 50)           1300        ['lstm[0][0]']

 dropout (Dropout)              (None, 50)           0           ['dense[0][0]']

 batch_normalization (BatchNorm  (None, 50)          200         ['dropout[0][0]']
 alization)

 dense_1 (Dense)                (None, 40)           2040        ['batch_normalization[0][0]']

 dropout_1 (Dropout)            (None, 40)           0           ['dense_1[0][0]']

 dense_2 (Dense)                (None, 10)           410         ['dropout_1[0][0]']

==================================================================================================
Total params: 6,650
Trainable params: 6,550
Non-trainable params: 100
__________________________________________________________________________________________________
```

```
In [30]: # https://stackoverflow.com/a/59564740

class AccThreshold(Callback):

    def __init__(self, thr_val, plus):
        self.thr_val = thr_val + plus

    def on_epoch_end(self, epoch, logs = {}):
        val_microF1 = logs.get('val_micro_f1')
        if (val_microF1 >= self.thr_val) and epoch >= 5:
            print(f'\n\n\tTerminating training at epoch {epoch+1} with a validation micro F1 score of {val_microF
            self.model.stop_training = True
```

```
In [31]: def call_back_list(thr_val, thr_plus, location):

    '''
    Generating keras Callbacks and return callback list

    Input
    -----
    1. Throshold value : Acceptance critiria
    2. Model stops at which accuracy
    3. Loaction to save ModelCheckPoint and Logs in model training
    '''

    cust_callback = AccThreshold(thr_val, thr_plus)

    model_pat  = f'results/Checkpoints/{location}/'
    filepath = model_pat + 'EPO_{epoch:02d}-F1_{val_micro_f1:.04}.h5'
    chek_pt = ModelCheckpoint(filepath, monitor = 'val_micro_f1', verbose = 0, save_best_only = True, )

    logDir = f'results/logs/{location}/' + datetime.now().strftime('%y_%h%d_%H%M')
    t_board = TensorBoard(log_dir = logDir, histogram_freq = 1)

    reduce_lr = ReduceLROnPlateau(monitor = 'val_micro_f1', factor = 0.2, verbose = 0, patience = 5)

    return [cust_callback, chek_pt, t_board, reduce_lr]
```

```
In [32]: def f1_micro(y_true,y_pred):
    return f1_score(y_true, y_pred, average = 'micro')

def micro_f1(y_true, y_proba):
    y_pred = tf.math.argmax(y_proba, axis = 1)
    return tf.py_function(f1_micro, (y_true, y_pred), tf.double)
```

```
In [33]: # https://keras.io/api/metrics/accuracy_metrics/#sparsecategoricalaccuracy-class
# https://keras.io/api/optimizers/adamax/
```

```
callBacks = call_back_list(0.1, 0, '1_Data_Raw')

model_raw.compile(optimizer = 'Adamax', loss = 'sparse_categorical_crossentropy', metrics = [micro_f1])

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

EPOCH = 40
raw_LSTM = model_raw.fit(x = [X_train_pad_seq, X_train_mask], y = y_train_int, epochs = EPOCH,
           validation_data = ([X_test_pad_seq, X_test_mask], y_test_int), callbacks = callBacks)
```

```
Epoch 1/40
44/44 [==============================] - 31s 506ms/step - loss: 2.3078 - micro_f1: 0.1001 - val_loss: 2.3026 - va
l_micro_f1: 0.1009 - lr: 0.0010
Epoch 2/40
44/44 [==============================] - 22s 511ms/step - loss: 2.3054 - micro_f1: 0.1030 - val_loss: 2.3026 - va
l_micro_f1: 0.1080 - lr: 0.0010
Epoch 3/40
44/44 [==============================] - 24s 548ms/step - loss: 2.3059 - micro_f1: 0.1006 - val_loss: 2.3026 - va
l_micro_f1: 0.1146 - lr: 0.0010
Epoch 4/40
44/44 [==============================] - 23s 516ms/step - loss: 2.3049 - micro_f1: 0.0940 - val_loss: 2.3026 - va
l_micro_f1: 0.0998 - lr: 0.0010
Epoch 5/40
44/44 [==============================] - 20s 462ms/step - loss: 2.3039 - micro_f1: 0.1058 - val_loss: 2.3026 - va
l_micro_f1: 0.0998 - lr: 0.0010
Epoch 6/40
44/44 [==============================] - 24s 548ms/step - loss: 2.3037 - micro_f1: 0.1155 - val_loss: 2.3026 - va
l_micro_f1: 0.0998 - lr: 0.0010
Epoch 7/40
44/44 [==============================] - 23s 521ms/step - loss: 2.3064 - micro_f1: 0.0895 - val_loss: 2.3026 - va
l_micro_f1: 0.0998 - lr: 0.0010
Epoch 8/40
44/44 [==============================] - 21s 471ms/step - loss: 2.3042 - micro_f1: 0.1103 - val_loss: 2.3026 - va
l_micro_f1: 0.0998 - lr: 0.0010
Epoch 9/40
44/44 [==============================] - 22s 508ms/step - loss: 2.3036 - micro_f1: 0.0933 - val_loss: 2.3026 - va
l_micro_f1: 0.0998 - lr: 0.0010
Epoch 10/40
44/44 [==============================] - 24s 551ms/step - loss: 2.3031 - micro_f1: 0.1042 - val_loss: 2.3026 - va
l_micro_f1: 0.0998 - lr: 2.0000e-04
Epoch 11/40
44/44 [==============================] - 23s 523ms/step - loss: 2.3024 - micro_f1: 0.0997 - val_loss: 2.3026 - va
l_micro_f1: 0.0998 - lr: 2.0000e-04
Epoch 12/40
44/44 [==============================] - ETA: 0s - loss: 2.3025 - micro_f1: 0.1134

        Terminating training at epoch 12 with a validation micro F1 score of 0.10143 %

44/44 [==============================] - 22s 506ms/step - loss: 2.3025 - micro_f1: 0.1134 - val_loss: 2.3026 - va
l_micro_f1: 0.1014 - lr: 2.0000e-04
```

In [34]:

```
def save_model_history(filename, base_model, model):

    '''
    Saves model and model history

    Input
    -----
    1. File name to store model history file
    2. Keras model name
    3. Saved keras model name
    '''

    with open('results/' + filename + '_history.pkl', 'wb') as file:
        pickle.dump(model.history, file)

    train_epo = len(model.epoch)
    final_f1 = model.history['val_micro_f1'][-1]
    base_model.save(f'results/{filename}_Epo_{train_epo}_F1_{final_f1:.4f}.h5')
```

In [35]:

```
save_model_history('1_data_raw', model_raw, raw_LSTM)
```

Observation

Using Raw data is not a good option because loss and micro-f1 is not imporving

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 [36]:

```
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 [37]:

```
# use convert_to_spectrogram and convert every raw sequence in X_train_pad_seq and X_test_pad-seq.
# save those all in the X_train_spectrogram and X_test_spectrogram ( These two arrays must be numpy arrays)

X_train_spectrogram  = np.array([convert_to_spectrogram(data) for data in tqdm(X_train_pad_seq, \
                                                                              'X_train_pad_seq : ')])

X_test_spectrogram  = np.array([convert_to_spectrogram(data) for data in tqdm(X_test_pad_seq, \
                                                                             'X_test_pad_seq : ')])
```

```
X_train_pad_seq : 100%|██████████| 1400/1400 [00:07<00:00, 183.78it/s]
X_test_pad_seq : 100%|██████████| 600/600 [00:03<00:00, 187.98it/s]
```

In [38]:

```
print(f'X_train_spectrogram --> Shape :: {X_train_spectrogram.shape}, Type :: {type(X_train_spectrogram)}')
print(f'X_test_spectrogram  --> Shape :: {X_test_spectrogram.shape} , Type :: {type(X_test_spectrogram)}')
```

```
X_train_spectrogram --> Shape :: (1400, 64, 35), Type :: <class 'numpy.ndarray'>
X_test_spectrogram  --> Shape :: (600, 64, 35) , Type :: <class 'numpy.ndarray'>
```

Grader function 6

In [39]:

```
def grader_spectrogram():

    flag_shape = (X_train_spectrogram.shape==(1400,64, 35)) and (X_test_spectrogram.shape == (600, 64, 35))
    return flag_shape

grader_spectrogram()
```

Out[39]:

```
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 [40]:

```
# write the architecture of the model
# print model.summary and make sure that it is following point 2 mentioned above

def model_2_4(max_length, model_name):

    tf.keras.backend.clear_session()

    input_ = Input(shape = X_train_spectrogram.shape[1:])
    x = LSTM(64, return_sequences = True)(input_)
    x = LSTM(64, return_sequences = True)(x)
```

```python
    x = GlobalAveragePooling1D()(x)
    x = Dense(50, activation = 'relu')(x)
    x = BatchNormalization()(x)
    x = Dense(20, activation = 'relu')(x)
    output_ = Dense(10, activation = 'softmax')(x)

    return Model(inputs = input_, outputs = output_, name = model_name)
```

In [41]:

```python
model_spectrogram = model_2_4(max_length, 'Spectrogram_Data_Alone')
model_spectrogram.summary()
```

```
Model: "Spectrogram_Data_Alone"
_________________________________________________________________
 Layer (type)                Output Shape              Param #   
=================================================================
 input_1 (InputLayer)        [(None, 64, 35)]          0         
                                                                 
 lstm (LSTM)                 (None, 64, 64)            25600     
                                                                 
 lstm_1 (LSTM)               (None, 64, 64)            33024     
                                                                 
 global_average_pooling1d (G  (None, 64)               0         
 lobalAveragePooling1D)                                          
                                                                 
 dense (Dense)               (None, 50)                3250      
                                                                 
 batch_normalization (BatchN  (None, 50)               200       
 ormalization)                                                   
                                                                 
 dense_1 (Dense)             (None, 20)                1020      
                                                                 
 dense_2 (Dense)             (None, 10)                210       
                                                                 
=================================================================
Total params: 63,304
Trainable params: 63,204
Non-trainable params: 100
_________________________________________________________________
```

In [42]:

```python
# By apply ReduceLROnPlateau it's not getting converged thus not applying it in this model

callBacks = call_back_list(0.8, 0.05, '2_Data_Spectrogram')[:-1]

model_spectrogram.compile(optimizer = 'Adamax', loss = 'sparse_categorical_crossentropy', metrics = [micro_f1])

# compile and fit your model.
# model2.fit([X_train_spectrogram],y_train_int,......)

EPOCH = 100

spectrogram_LSTM = model_spectrogram.fit(x = X_train_spectrogram, y = y_train_int, epochs = EPOCH,
                      validation_data = (X_test_spectrogram, y_test_int), callbacks = callBacks)
```

```
Epoch 1/100
 4/44 [=>............................] - ETA: 0s - loss: 2.3369 - micro_f1: 0.1172
WARNING:tensorflow:Callback method `on_train_batch_end` is slow compared to the batch time (batch time: 0.0148s v
s `on_train_batch_end` time: 0.0283s). Check your callbacks.
44/44 [==============================] - 6s 58ms/step - loss: 2.1658 - micro_f1: 0.2150 - val_loss: 2.2818 - val_
micro_f1: 0.1036
Epoch 2/100
44/44 [==============================] - 1s 26ms/step - loss: 1.9986 - micro_f1: 0.3182 - val_loss: 2.2526 - val_
micro_f1: 0.1464
Epoch 3/100
44/44 [==============================] - 1s 18ms/step - loss: 1.8613 - micro_f1: 0.3568 - val_loss: 2.2120 - val_
micro_f1: 0.2182
Epoch 4/100
44/44 [==============================] - 1s 16ms/step - loss: 1.7404 - micro_f1: 0.4034 - val_loss: 2.1051 - val_
micro_f1: 0.3273
Epoch 5/100
44/44 [==============================] - 1s 17ms/step - loss: 1.6198 - micro_f1: 0.4531 - val_loss: 1.9839 - val_
micro_f1: 0.4331
Epoch 6/100
44/44 [==============================] - 1s 17ms/step - loss: 1.5029 - micro_f1: 0.4953 - val_loss: 1.8805 - val_
micro_f1: 0.4567
Epoch 7/100
44/44 [==============================] - 1s 16ms/step - loss: 1.4011 - micro_f1: 0.5320 - val_loss: 1.7475 - val_
micro_f1: 0.4194
Epoch 8/100
44/44 [==============================] - 1s 17ms/step - loss: 1.3364 - micro_f1: 0.5639 - val_loss: 1.5717 - val_
micro_f1: 0.5461
Epoch 9/100
44/44 [==============================] - 1s 17ms/step - loss: 1.2663 - micro_f1: 0.5784 - val_loss: 1.6804 - val_
micro_f1: 0.4117
```

```
Epoch 10/100
44/44 [==============================] - 1s 17ms/step - loss: 1.2417 - micro_f1: 0.5926 - val_loss: 1.3776 - val_
micro_f1: 0.5570
Epoch 11/100
44/44 [==============================] - 1s 17ms/step - loss: 1.1377 - micro_f1: 0.6314 - val_loss: 1.2788 - val_
micro_f1: 0.6245
Epoch 12/100
44/44 [==============================] - 1s 17ms/step - loss: 1.0795 - micro_f1: 0.6503 - val_loss: 1.2320 - val_
micro_f1: 0.6239
Epoch 13/100
44/44 [==============================] - 1s 18ms/step - loss: 1.0277 - micro_f1: 0.6731 - val_loss: 1.2193 - val_
micro_f1: 0.5746
Epoch 14/100
44/44 [==============================] - 1s 18ms/step - loss: 0.9727 - micro_f1: 0.6906 - val_loss: 1.1693 - val_
micro_f1: 0.6184
Epoch 15/100
44/44 [==============================] - 1s 17ms/step - loss: 0.8942 - micro_f1: 0.7157 - val_loss: 1.0349 - val_
micro_f1: 0.6579
Epoch 16/100
44/44 [==============================] - 1s 17ms/step - loss: 0.8574 - micro_f1: 0.7247 - val_loss: 1.0377 - val_
micro_f1: 0.6530
Epoch 17/100
44/44 [==============================] - 1s 17ms/step - loss: 0.8164 - micro_f1: 0.7332 - val_loss: 0.9322 - val_
micro_f1: 0.6826
Epoch 18/100
44/44 [==============================] - 1s 17ms/step - loss: 0.7837 - micro_f1: 0.7493 - val_loss: 0.9183 - val_
micro_f1: 0.6979
Epoch 19/100
44/44 [==============================] - 1s 16ms/step - loss: 0.7630 - micro_f1: 0.7545 - val_loss: 1.0552 - val_
micro_f1: 0.6409
Epoch 20/100
44/44 [==============================] - 1s 17ms/step - loss: 0.7158 - micro_f1: 0.7794 - val_loss: 0.9087 - val_
micro_f1: 0.7039
Epoch 21/100
44/44 [==============================] - 1s 18ms/step - loss: 0.6638 - micro_f1: 0.7933 - val_loss: 0.9289 - val_
micro_f1: 0.6711
Epoch 22/100
44/44 [==============================] - 1s 17ms/step - loss: 0.6284 - micro_f1: 0.7957 - val_loss: 0.7719 - val_
micro_f1: 0.7615
Epoch 23/100
44/44 [==============================] - 1s 17ms/step - loss: 0.6295 - micro_f1: 0.7976 - val_loss: 0.8384 - val_
micro_f1: 0.7149
Epoch 24/100
44/44 [==============================] - 1s 18ms/step - loss: 0.6016 - micro_f1: 0.7981 - val_loss: 0.9290 - val_
micro_f1: 0.7039
Epoch 25/100
44/44 [==============================] - 1s 16ms/step - loss: 0.5483 - micro_f1: 0.8305 - val_loss: 0.7336 - val_
micro_f1: 0.7577
Epoch 26/100
44/44 [==============================] - 1s 16ms/step - loss: 0.5099 - micro_f1: 0.8411 - val_loss: 0.7225 - val_
micro_f1: 0.7719
Epoch 27/100
44/44 [==============================] - 1s 17ms/step - loss: 0.5115 - micro_f1: 0.8397 - val_loss: 0.8893 - val_
micro_f1: 0.7007
Epoch 28/100
44/44 [==============================] - 1s 17ms/step - loss: 0.4814 - micro_f1: 0.8452 - val_loss: 0.7137 - val_
micro_f1: 0.7763
Epoch 29/100
44/44 [==============================] - 1s 17ms/step - loss: 0.4593 - micro_f1: 0.8651 - val_loss: 0.7943 - val_
micro_f1: 0.7286
Epoch 30/100
44/44 [==============================] - 1s 17ms/step - loss: 0.4162 - micro_f1: 0.8771 - val_loss: 0.6640 - val_
micro_f1: 0.7818
Epoch 31/100
44/44 [==============================] - 1s 17ms/step - loss: 0.4124 - micro_f1: 0.8864 - val_loss: 0.6997 - val_
micro_f1: 0.7577
Epoch 32/100
44/44 [==============================] - 1s 16ms/step - loss: 0.3896 - micro_f1: 0.8707 - val_loss: 0.7678 - val_
micro_f1: 0.7615
Epoch 33/100
44/44 [==============================] - 1s 17ms/step - loss: 0.3707 - micro_f1: 0.8821 - val_loss: 0.6130 - val_
micro_f1: 0.8169
Epoch 34/100
44/44 [==============================] - 1s 17ms/step - loss: 0.3449 - micro_f1: 0.8911 - val_loss: 0.6699 - val_
micro_f1: 0.7664
Epoch 35/100
44/44 [==============================] - 1s 17ms/step - loss: 0.3204 - micro_f1: 0.9110 - val_loss: 0.8060 - val_
micro_f1: 0.7533
Epoch 36/100
44/44 [==============================] - 1s 17ms/step - loss: 0.3357 - micro_f1: 0.8920 - val_loss: 0.5979 - val_
micro_f1: 0.8043
Epoch 37/100
44/44 [==============================] - 1s 18ms/step - loss: 0.3165 - micro_f1: 0.9105 - val_loss: 0.6229 - val_
micro_f1: 0.7895
Epoch 38/100
41/44 [==========================>...] - ETA: 0s - loss: 0.3065 - micro_f1: 0.9047

        Terminating training at epoch 38 with a validation micro F1 score of 0.85033 %
```

```
44/44 [==============================] - 1s 17ms/step - loss: 0.3108 - micro_f1: 0.9032 - val_loss: 0.5027 - val_
micro_f1: 0.8503
```

In [43]:
```
save_model_history('2_data_spectrogram', model_spectrogram, spectrogram_LSTM)
```

Observation

Using spectogram data the micro_f1 score increased drastically than raw data model.

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 [44]:
```
# 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, n_steps=pitch_value)
            augmented_data.append(final_data)

    return augmented_data
```

In [45]:
```
temp_path = df_audio.iloc[0].path
aug_temp = generate_augmented_data(temp_path)
```

In [46]:
```
len(aug_temp)
```

Out[46]: 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 [47]:
```
X_train, X_test, y_train, y_test = train_test_split(df_audio['path'], df_audio['label'],
                                        random_state = 45, test_size = 0.2, stratify = df_audio['label'])
```

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

In [48]:
```
# https://stackoverflow.com/a/41175538

def get_augmented_n_idx_data(df, name_):

    aug_data = []
    idx_values = []

    for idx, data in tqdm(zip(df.index, df), total = len(df), desc = name_):
```

```
        augment_values = generate_augmented_data(data)
        for aug in augment_values:
            aug_data.append(aug)
            idx_values.append(idx)

    return aug_data, idx_values

def get_augmented_data_for_y(df, idx_list, name_):

    aug_labels = []

    for idx, label in tqdm(zip(df.index, df), total = len(df), desc = name_):
        for idx_val in idx_list:
            if idx == idx_val:
                aug_labels.append(label)

    return aug_labels
```

In [49]:

```
augmented_train_data, idx_values_train = get_augmented_n_idx_data(X_train, 'X_train Aug. ')
augment_y_train = get_augmented_data_for_y(y_train, idx_values_train, 'y_train Aug. ')
```

```
X_train Aug. : 100%|██████████| 1600/1600 [04:50<00:00,  5.50it/s]
y_train Aug. : 100%|██████████| 1600/1600 [00:00<00:00, 2250.54it/s]
```

In [50]:

```
print(f"Length of 'augmented_train_data' :: {len(augmented_train_data)}")
print(f"Length of 'augment_y_train' :: {len(augment_y_train)}")
```

```
Length of 'augmented_train_data' :: 14400
Length of 'augment_y_train' :: 14400
```

In [51]:

```
# Using only raw_data NOT duration from the process_wave_data() output

test_data = process_wave_data(X_test, 'X_test').raw_data
```

```
X_test : 100%|██████████| 400/400 [00:03<00:00, 106.98it/s]
```

In [52]:

```
augmented_train_data = np.array(augmented_train_data)
print(f"Type of 'augmented_train_data' :: {type(augmented_train_data)}")

test_data = np.array(test_data)
print(f"Type of 'test_data_processed'  :: {type(test_data)}")

print(f"Type of 'augment_y_train'  :: {type(augment_y_train)}")
print("\nConverting 'augment_y_train' from 'Python list' to 'Numpy Array'")

augment_y_train = np.array(augment_y_train)

print(f"\nType of 'augment_y_train'  :: {type(augment_y_train)}")

aug_y_train_int = augment_y_train.astype('int')
y_test_int = y_test.astype('int')
```

```
Type of 'augmented_train_data' :: <class 'numpy.ndarray'>
Type of 'test_data_processed'  :: <class 'numpy.ndarray'>
Type of 'augment_y_train'  :: <class 'list'>

Converting 'augment_y_train' from 'Python list' to 'Numpy Array'

Type of 'augment_y_train'  :: <class 'numpy.ndarray'>
```

In [53]:

```
# https://stackoverflow.com/a/63853924

max_length  = 17640

augmented_train_data_seq = pad_sequences(augmented_train_data, maxlen = max_length, padding = 'post',
                                                     value = 0, dtype = 'float', truncating = 'post')

test_data_seq = pad_sequences(test_data, maxlen = max_length, padding = 'post', value = 0,
                                                      dtype = 'float', truncating = 'post')

print(f"Shape of 'augmented_train_data_seq' :: {augmented_train_data_seq.shape}")
print(f"Shape of 'test_data_seq' :: {test_data_seq.shape}")
```

```
print(f"\nShape of 'aug_y_train_int' :: {aug_y_train_int.shape}")
print(f"Shape of 'y_test_int' :: {y_test_int.shape}")

train_augmented_seq_mask = augmented_train_data_seq.astype('bool')
test_data_seq_mask = test_data_seq.astype('bool')
```

```
Shape of 'augmented_train_data_seq' :: (14400, 17640)
Shape of 'test_data_seq' :: (400, 17640)

Shape of 'aug_y_train_int' :: (14400,)
Shape of 'y_test_int' :: (400,)
```

In [54]:
```
model_augmented = model_1_3(max_length, 'Data_Augmented')

model_augmented.summary()
```

```
Model: "Data_Augmented"
__________________________________________________________________________________________________
 Layer (type)                   Output Shape         Param #     Connected to                     
==================================================================================================
 input_1 (InputLayer)           [(None, 17640, 1)]   0           []                               

 input_2 (InputLayer)           [(None, 17640)]      0           []                               

 lstm (LSTM)                    (None, 25)           2700        ['input_1[0][0]',                
                                                                  'input_2[0][0]']                

 dense (Dense)                  (None, 50)           1300        ['lstm[0][0]']                   

 dropout (Dropout)              (None, 50)           0           ['dense[0][0]']                  

 batch_normalization (BatchNorm  (None, 50)          200         ['dropout[0][0]']                
 alization)                                                                                       

 dense_1 (Dense)                (None, 40)           2040        ['batch_normalization[0][0]']    

 dropout_1 (Dropout)            (None, 40)           0           ['dense_1[0][0]']                

 dense_2 (Dense)                (None, 10)           410         ['dropout_1[0][0]']              

==================================================================================================
Total params: 6,650
Trainable params: 6,550
Non-trainable params: 100
__________________________________________________________________________________________________
```

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 [55]:
```
callBacks = call_back_list(0.1, 0.003, '3_Data_Augmented')

model_augmented.compile(optimizer = 'Adamax', loss = 'sparse_categorical_crossentropy', metrics = [micro_f1])

EPOCH = 40

augmented_LSTM = model_augmented.fit(x = [augmented_train_data_seq, train_augmented_seq_mask],
                                     validation_data = ([test_data_seq, test_data_seq_mask], y_test_int),
                                     y = aug_y_train_int, epochs = EPOCH, callbacks = callBacks)
```

```
Epoch 1/40
450/450 [==============================] - 197s 427ms/step - loss: 2.3038 - micro_f1: 0.1020 - val_loss: 2.3026 - val_micro_f1: 0.0962 - lr: 0.0010
Epoch 2/40
450/450 [==============================] - 195s 434ms/step - loss: 2.3028 - micro_f1: 0.1010 - val_loss: 2.3029 - val_micro_f1: 0.1034 - lr: 0.0010
Epoch 3/40
450/450 [==============================] - 195s 433ms/step - loss: 2.3029 - micro_f1: 0.0988 - val_loss: 2.3025 - val_micro_f1: 0.0817 - lr: 0.0010
Epoch 4/40
450/450 [==============================] - 196s 436ms/step - loss: 2.3028 - micro_f1: 0.0999 - val_loss: 2.3026 - val_micro_f1: 0.0841 - lr: 0.0010
Epoch 5/40
450/450 [==============================] - 194s 432ms/step - loss: 2.3029 - micro_f1: 0.0937 - val_loss: 2.3027 - val_micro_f1: 0.1058 - lr: 0.0010
```

```
Epoch 6/40
450/450 [==============================] - 202s 448ms/step - loss: 2.3027 - micro_f1: 0.0946 - val_loss: 2.3026 - val_micro_f1: 0.0962 - lr: 0.0010
Epoch 7/40
450/450 [==============================] - ETA: 0s - loss: 2.3027 - micro_f1: 0.0957

        Terminating training at epoch 7 with a validation micro F1 score of 0.10337 %

450/450 [==============================] - 195s 434ms/step - loss: 2.3027 - micro_f1: 0.0957 - val_loss: 2.3025 - val_micro_f1: 0.1034 - lr: 0.0010
```

In [56]:

```python
save_model_history('3_data_augmented', model_augmented, augmented_LSTM)
```

Observation

Using augmented data the the loss became loss has reduced.

4. Data augmentation with spectogram data

1. use convert_to_spectrogram and convert the padded data from train and test data to spectogram data.
2. The shape of train data will be 14400 x 64 x 35 and shape of test_data will be 400 x 64 x35
3. Define the model similar to model 2 and fit the data

In [57]:

```python
# Using augmented_train_data_seq, y_train_int, test_data_seq and y_test_int from model_3

augmented_train_data_spectrogram  = [convert_to_spectrogram(data) for data in tqdm(augmented_train_data_seq,
                                                                                     'X_train_pad_seq : ')]
test_data_spectrogram  = [convert_to_spectrogram(data) for data in tqdm(test_data_seq, 'X_test_pad_seq : ')]

augmented_train_data_spectrogram = np.array(augmented_train_data_spectrogram)
test_data_spectrogram = np.array(test_data_spectrogram)
```

```
X_train_pad_seq : 100%|██████████| 14400/14400 [06:09<00:00, 38.98it/s]
X_test_pad_seq : 100%|██████████| 400/400 [00:11<00:00, 35.07it/s]
```

In [58]:

```python
print(f"Shape of 'augmented_train_data_spectrogram' :: {augmented_train_data_spectrogram.shape}")
print(f"Shape of 'test_data_spectrogram' :: {test_data_spectrogram.shape}")

print(f"\nShape of 'aug_y_train_int' :: {aug_y_train_int.shape}")
print(f"Shape of 'y_test_int' :: {y_test_int.shape}")
```

```
Shape of 'augmented_train_data_spectrogram' :: (14400, 64, 35)
Shape of 'test_data_spectrogram' :: (400, 64, 35)

Shape of 'aug_y_train_int' :: (14400,)
Shape of 'y_test_int' :: (400,)
```

In [59]:

```python
model_aug_spectro = model_2_4(max_length, 'Data_Augment_nd_Spectro')

model_aug_spectro.summary()
```

```
Model: "Data_Augment_nd_Spectro"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 64, 35)]          0

 lstm (LSTM)                 (None, 64, 64)            25600

 lstm_1 (LSTM)               (None, 64, 64)            33024

 global_average_pooling1d (G  (None, 64)               0
 lobalAveragePooling1D)

 dense (Dense)               (None, 50)                3250

 batch_normalization (BatchN  (None, 50)               200
 ormalization)

 dense_1 (Dense)             (None, 20)                1020

 dense_2 (Dense)             (None, 10)                210
```

```
=================================================================
Total params: 63,304
Trainable params: 63,204
Non-trainable params: 100
_________________________________________________________________
```

In [60]:

```python
callBacks = call_back_list(0.8, 0.05, '4_Data_Augment_Spectro') #[:-1] # Not applying ReduceLROnPlateau callback

model_aug_spectro.compile(optimizer = 'Adamax', loss = 'sparse_categorical_crossentropy', metrics = [micro_f1])

EPOCH = 100

aug_spectro_LSTM = model_aug_spectro.fit(x = augmented_train_data_spectrogram, y = aug_y_train_int,
                                        validation_data = (test_data_spectrogram, y_test_int),
                                        epochs = EPOCH,  callbacks = callBacks)
```

```
Epoch 1/100
  6/450 [..............................] - ETA: 5s - loss: 2.3390 - micro_f1: 0.1354
WARNING:tensorflow:Callback method `on_train_batch_end` is slow compared to the batch time (batch time: 0.0090s v
s `on_train_batch_end` time: 0.0182s). Check your callbacks.
450/450 [==============================] - 8s 12ms/step - loss: 1.8926 - micro_f1: 0.3207 - val_loss: 1.6771 - va
l_micro_f1: 0.4471 - lr: 0.0010
Epoch 2/100
450/450 [==============================] - 5s 10ms/step - loss: 1.3567 - micro_f1: 0.5260 - val_loss: 1.0475 - va
l_micro_f1: 0.6298 - lr: 0.0010
Epoch 3/100
450/450 [==============================] - 5s 10ms/step - loss: 0.9787 - micro_f1: 0.6631 - val_loss: 0.8653 - va
l_micro_f1: 0.7308 - lr: 0.0010
Epoch 4/100
450/450 [==============================] - 5s 10ms/step - loss: 0.7332 - micro_f1: 0.7448 - val_loss: 0.8028 - va
l_micro_f1: 0.7091 - lr: 0.0010
Epoch 5/100
450/450 [==============================] - 5s 10ms/step - loss: 0.5696 - micro_f1: 0.8056 - val_loss: 0.5877 - va
l_micro_f1: 0.8245 - lr: 0.0010
Epoch 6/100
450/450 [==============================] - 5s 11ms/step - loss: 0.4626 - micro_f1: 0.8416 - val_loss: 0.6320 - va
l_micro_f1: 0.7909 - lr: 0.0010
Epoch 7/100
449/450 [============================>.] - ETA: 0s - loss: 0.3476 - micro_f1: 0.8847

        Terminating training at epoch 7 with a validation micro F1 score of 0.88462 %

450/450 [==============================] - 6s 13ms/step - loss: 0.3475 - micro_f1: 0.8847 - val_loss: 0.3955 - va
l_micro_f1: 0.8846 - lr: 2.0000e-04
```

In [61]:

```python
save_model_history('4_data_aug_spectro', model_aug_spectro, aug_spectro_LSTM)
```

Observation

Using data augmentation and spectogram the micro_f1 score increased drastically than spectrogram alone data model and also helped to reduce loss and converged in much more faster.

In [62]:

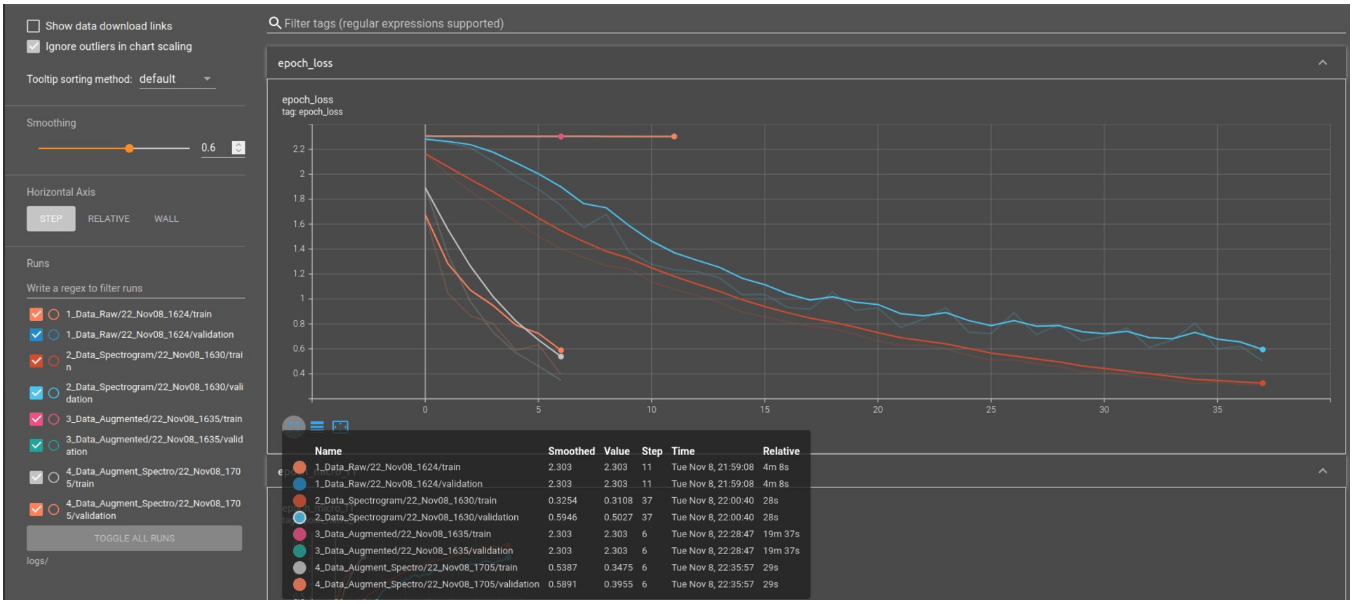
```python
model_list = [('RAW_Data_Alone', raw_LSTM), ('Spectrogram_Data_Alone', spectrogram_LSTM),
              ('Data_Augmented', augmented_LSTM), ('Data_Augment_nd_Spectro', aug_spectro_LSTM)]

table = PrettyTable(['Sl.No', 'Model Name', 'Epochs Ran', 'Micro-F1', 'Val. Micro-F1',
                     'Change in Loss (initial-final)'], hrules = True)

table.set_style(SINGLE_BORDER)

for idx, (name, model)  in enumerate(model_list):

    model = model.history
    epochs = len(model['val_loss'])
    micro_f1 = round(model['micro_f1'][-1], 5)
    val_micro_f1 = round(model['val_micro_f1'][-1], 5)
    loss = round(model['loss'][0] - model['loss'][-1], 5)

    table.add_row([idx+1, name, epochs, micro_f1, val_micro_f1, loss])

print(table)
```

```
┌───────┬─────────────────────────┬────────────┬──────────┬───────────────┬────────────────────────────────┐
│ Sl.No │        Model Name       │ Epochs Ran │ Micro-F1 │ Val. Micro-F1 │ Change in Loss (initial-final) │
```

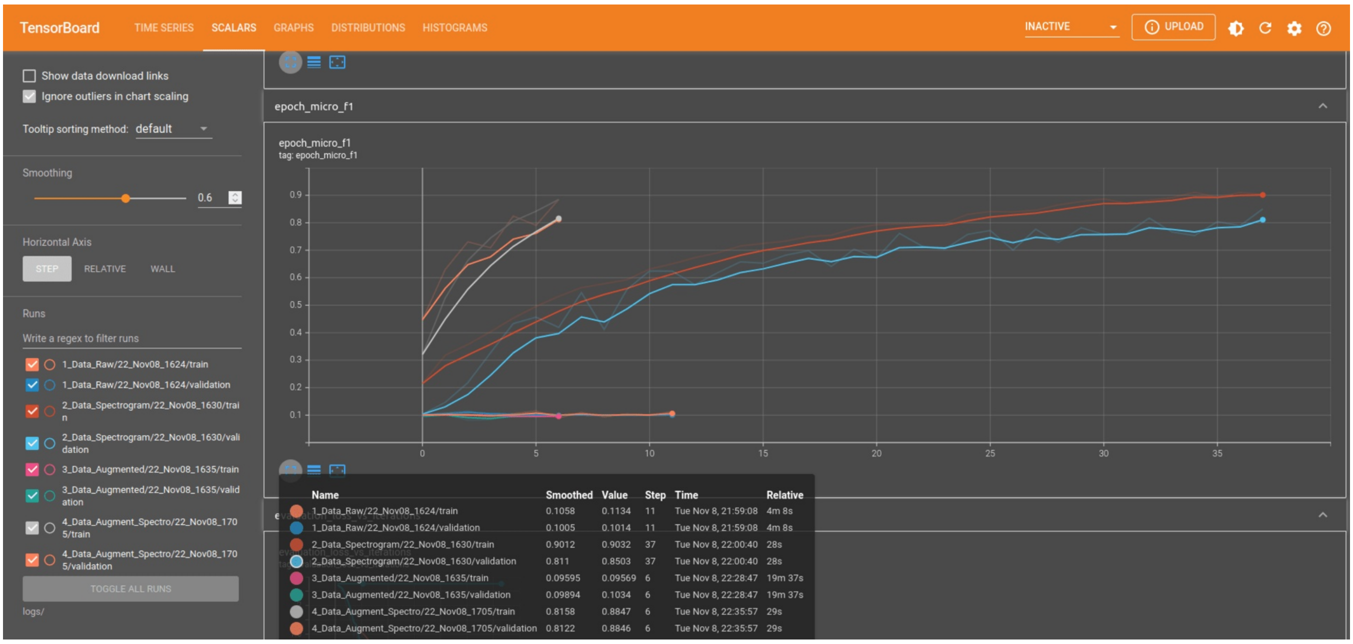
1	RAW_Data_Alone	12	0.1134	0.10143	0.00528
2	Spectrogram_Data_Alone	38	0.90317	0.85033	1.85505
3	Data_Augmented	7	0.09569	0.10337	0.00112
4	Data_Augment_nd_Spectro	7	0.88472	0.88462	1.54503

Tensorboard Results

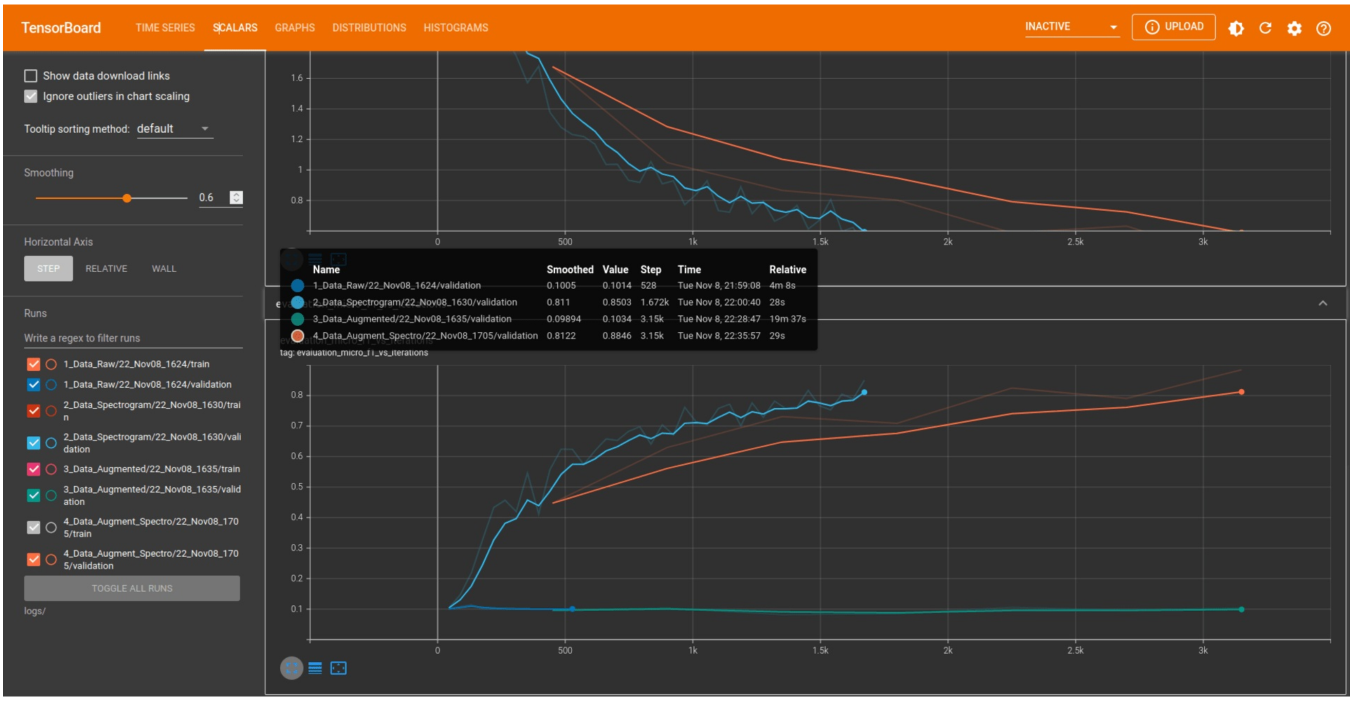
Epoch vs Loss



Epoch vs Micro_F1



Evaluation Micro_F1 vs Iterations



Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js