```
import librosa
import soundfile
import os, glob, pickle
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import accuracy_score


#Extracting features from audio files
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        # STFT
        stft=np.abs(librosa.stft(X))
        # MFCC
        mfccs=np.mean(librosa.feature.mfcc(y=X, sr=sample_rate, n_mfcc=50).T, axis=0)
        result=np.hstack((result, mfccs))
        return result


#Emotions present in the RAVDESS dataset
emotions={
  '01':'neutral',
  '02':'calm',
  '03':'happy',
  '04':'sad',
  '05':'angry',
  '06':'fearful',
  '07':'disgust',
  '08':'surprised'
}


# Emotions to recognize
emotions_to_recognize=['angry','disgust','surprised','calm','neutral','happy','sad','fearful']


# Data Loading and feature extraction for each sound file
def load_data(test_size=0.2):
    x,y=[],[]
    for file in glob.glob('/content/drive/MyDrive/RAVDESS/**/*.wav'):
        file_name=os.path.basename(file)
        emotion=emotions[file_name.split("-")[2]]
        if emotion not in emotions_to_recognize:
            continue
        feature=feature_extraction(file)
        x.append(feature)
        y.append(emotion)
    return train_test_split(np.array(x), y, test_size=test_size, random_state=9)


# Dataset splitting for training and testing
x_train,x_test,y_train,y_test= load_data(test_size=0.2)


# training and testing datasets
print((x_train.shape[0], x_test.shape[0]))
```

```
(1152, 288)
```

```
# the features extracted
print(f'Features extracted: {x_train.shape[1]}')
```

```
Features extracted: 40
```

```
# Model Initialization -> The Multi Layer Perceptron Classifier
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)


# Model Training
model.fit(x_train,y_train)
```

```
MLPClassifier(alpha=0.01, batch_size=256, hidden_layer_sizes=(300,),
              learning_rate='adaptive', max_iter=500)
```

```
# Model Prediction
y_pred=model.predict(x_test)
```

```
# Model Accuracy
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
Accuracy: 48.61%
```

```
x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 50
Accuracy: 50.00%
```

Assessing the accuracy of the model with each feature independently

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        # STFT
        stft=np.abs(librosa.stft(X))
        # Chroma_STFT
        chroma=np.mean(librosa.feature.chroma_stft(S=stft, sr=sample_rate).T,axis=0)
        result=np.hstack((result, chroma))
        return result
```

```
x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 12
Accuracy: 16.67%
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        # Mel Spectogram
        mel=np.mean(librosa.feature.melspectrogram(X, sr=sample_rate).T,axis=0)
        result=np.hstack((result, mel))
        return result
```

```
x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 128
Accuracy: 40.28%
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        zcr = np.mean(librosa.feature.zero_crossing_rate(y=X).T, axis=0)
        result=np.hstack((result, zcr))
        return result
```

```
x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 1
Accuracy: 20.14%
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        cqt = np.mean(librosa.feature.chroma_cqt(y=X, sr=sample_rate).T, axis=0)
        result = np.hstack((result, cqt))
        return result
```

```
x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 12
Accuracy: 26.74%
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        chroma_cens = np.mean(librosa.feature.chroma_cens(y=X, sr=sample_rate).T, axis=0)
        result = np.hstack((result, chroma_cens))
        return result
```

```
x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 12
Accuracy: 26.74%
/usr/local/lib/python3.8/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:692: ConvergenceWarning: Stochastic Optimiz
  warnings.warn(
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        stft=np.abs(librosa.stft(X))
        chroma_spectral_poly_features = np.mean(librosa.feature.poly_features(S=stft, order=2).T, axis=0)
        result = np.hstack((result, chroma_spectral_poly_features))
        return result
```

```
x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 3
Accuracy: 30.21%
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        stft=np.abs(librosa.stft(X))
        chroma_spectral_centroid = np.mean(librosa.feature.spectral_centroid(S=stft, sr=sample_rate).T, axis=0)
        result = np.hstack((result, chroma_spectral_centroid))
        return result


x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 1
Accuracy: 12.50%
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        stft=np.abs(librosa.stft(X))
        chroma_spectral_bandwidth = np.mean(librosa.feature.spectral_bandwidth(S=stft, sr=sample_rate).T, axis=0)
        result = np.hstack((result, chroma_spectral_bandwidth))
        return result


x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 1
Accuracy: 12.50%
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        stft=np.abs(librosa.stft(X))
        chroma_spectral_contrast = np.mean(librosa.feature.spectral_contrast(S=stft, sr=sample_rate).T, axis=0)
        result = np.hstack((result, chroma_spectral_contrast))
        return result


x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 7
Accuracy: 23.26%
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
```

```
        stft=np.abs(librosa.stft(X))
        chroma_spectral_flatness = np.mean(librosa.feature.spectral_flatness(S=stft).T, axis=0)
        result = np.hstack((result, chroma_spectral_flatness))
        return result

x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 1
Accuracy: 11.11%
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        stft=np.abs(librosa.stft(X))
        chroma_spectral_rolloff = np.mean(librosa.feature.spectral_rolloff(S=stft, sr=sample_rate).T, axis=0)
        result = np.hstack((result, chroma_spectral_rolloff))
        return result


x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 1
Accuracy: 12.50%
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        rms = np.mean(librosa.feature.rms(y=X).T, axis=0)
        result = np.hstack((result, rms))
        return result


x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 1
Accuracy: 27.43%
/usr/local/lib/python3.8/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:692: ConvergenceWarning: Stochastic Optimiz
  warnings.warn(
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        harmonic = np.abs(librosa.effects.harmonic(X))
        tonnetz = np.mean(librosa.feature.tonnetz(y=harmonic, sr=sample_rate).T, axis=0)
        result = np.hstack((result, tonnetz))
        return result


x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
```

```
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 6
Accuracy: 15.97%
/usr/local/lib/python3.8/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:692: ConvergenceWarning: Stochastic Optimiz
  warnings.warn(
```

```
def feature_extraction(file_name):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        result=np.array([])
        cqt = np.mean(librosa.feature.chroma_cqt(y=X, sr=sample_rate).T, axis=0)
        result = np.hstack((result, cqt))
        return result


x_train,x_test,y_train,y_test= load_data(test_size=0.2)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
```

```
(1152, 288)
Features extracted: 12
Accuracy: 27.08%
/usr/local/lib/python3.8/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:692: ConvergenceWarning: Stochastic Optimiz
  warnings.warn(
```

Now we have data on how well each feature can help in our SER Tonnetz: 15.97%, Root Mean Square: 27.43%, chroma_spectral_centroid: 12.5%, chroma_spectral_bandwidth: 12.5%, chroma_spectral_contrast: 23.26%, chroma_spectral_flatness: 11.11%, chroma_spectral_rolloff: 12.5%, chroma_spectral_poly_features: 30.21% (Order: 2), chroma_CENS: 26.74%, CQT: 27.08 %, Zero Crossing Data: 20.14%, Mel spectrogram: 40.28%, MFCC: 48.61%, Chroma_STFT: 16.67%

We can start experimenting different combinations of these features and see which of these when fused improves accuracy