

urban-sound-analysis

January 31, 2024

0.1 Dataset Information

This dataset contains 8732 labeled sound excerpts (≤ 4 s) of urban sounds from 10 classes:

air_conditioner
car_horn
children_playing
dog_bark
drilling
engine_idling
gun_shot
jackhammer
siren
street_music

0.2 Mounting Drive

```
[1]: from google.colab import drive  
drive.mount('/content/drive')
```

Mounted at /content/drive

```
[2]: !pwd
```

/content

0.3 Unzip data

```
[1]: !unzip 'drive/MyDrive/Colab Notebooks/train.zip'
```

0.4 Import modules

```
[16]: import pandas as pd
import numpy as np
import librosa
import librosa.display
import glob
import IPython.display as ipd
import random
%pylab inline

import warnings
warnings.filterwarnings('ignore')
```

Populating the interactive namespace from numpy and matplotlib

0.5 Loading the dataset

```
[6]: df = pd.read_csv('Urban Sound Dataset.csv')
df.head()
```

```
[6]:   ID      Class
0   0      siren
1   1 street_music
2   2    drilling
3   3      siren
4   4    dog_bark
```

```
[9]: ipd.Audio('Train/1.wav')
```

```
[9]: <IPython.lib.display.Audio object>
```

0.6 Exploratory Data Analysis

```
[10]: data, sampling_rate = librosa.load('Train/1.wav')
```

```
[11]: data
```

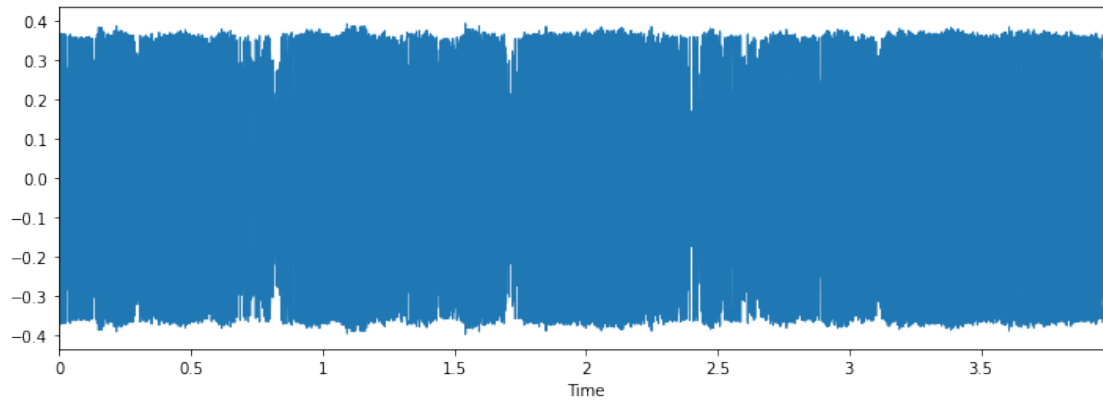
```
[11]: array([-0.09602016, -0.14303702,  0.05203498, ..., -0.01646687,
        -0.00915894,  0.09742922], dtype=float32)
```

```
[12]: sampling_rate
```

```
[12]: 22050
```

```
[15]: plt.figure(figsize=(12,4))
librosa.display.waveplot(data, sr=sampling_rate)
```

[15]: <matplotlib.collections.PolyCollection at 0x7f135da27550>



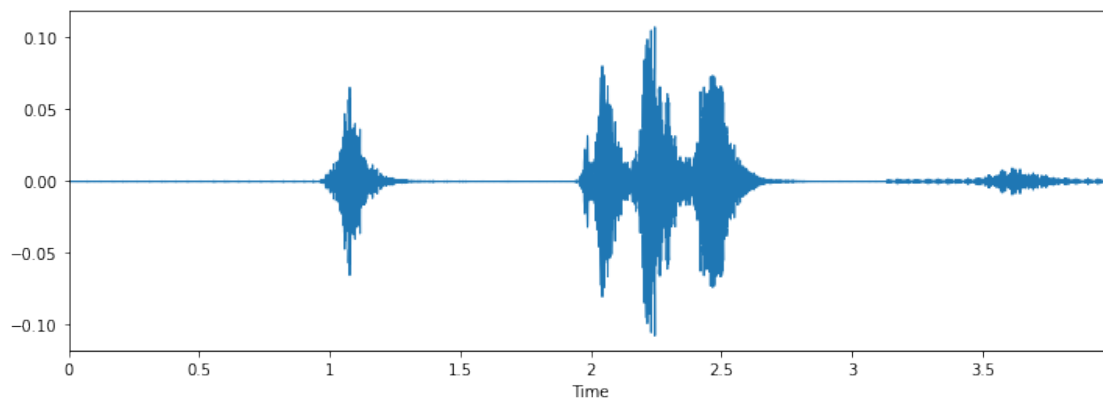
```
[28]: index = random.choice(df.index)

print('Class:', df['Class'][index])
data, sampling_rate = librosa.load('Train/'+str(df['ID'][index]) + '.wav')

plt.figure(figsize=(12,4))
librosa.display.waveplot(data, sr=sampling_rate)
```

Class: dog_bark

[28]: <matplotlib.collections.PolyCollection at 0x7f1354f76160>



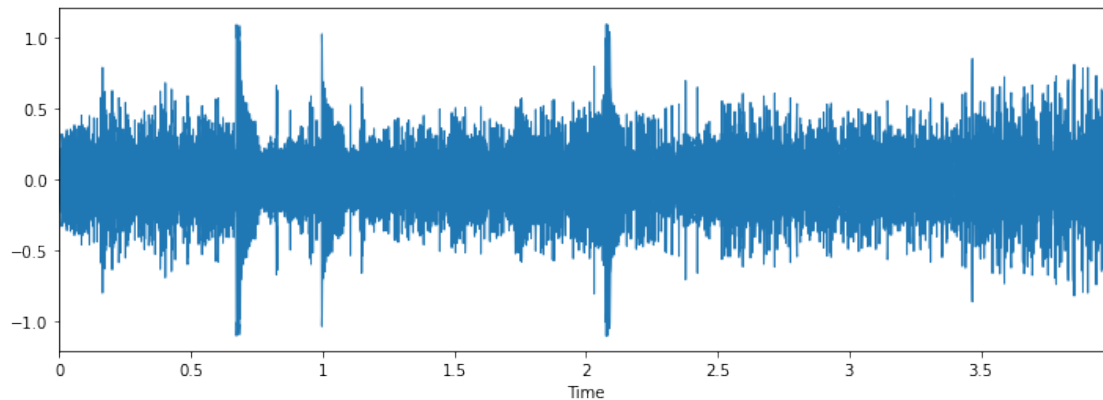
```
[29]: index = random.choice(df.index)

print('Class:', df['Class'][index])
data, sampling_rate = librosa.load('Train/'+str(df['ID'][index]) + '.wav')
```

```
plt.figure(figsize=(12,4))
librosa.display.waveplot(data, sr=sampling_rate)
```

Class: gun_shot

[29]: <matplotlib.collections.PolyCollection at 0x7f1354f570f0>



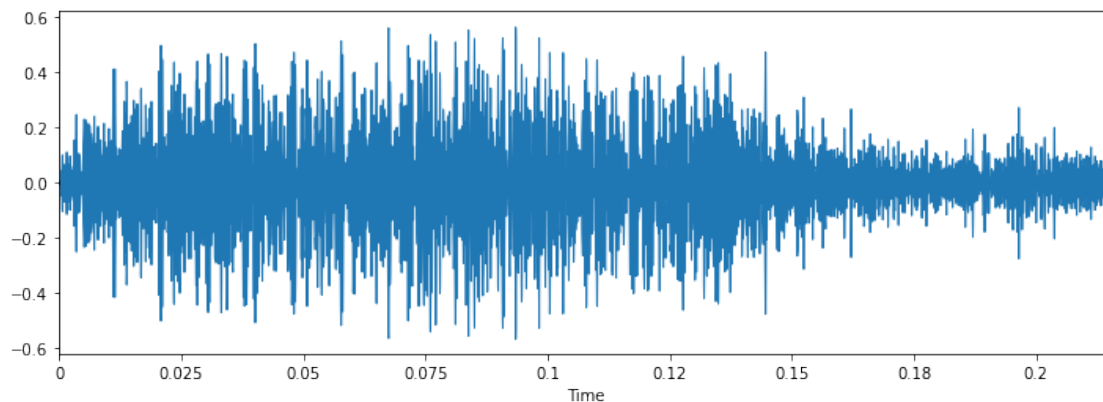
```
[30]: index = random.choice(df.index)

print('Class:', df['Class'][index])
data, sampling_rate = librosa.load('Train/'+str(df['ID'][index]) + '.wav')

plt.figure(figsize=(12,4))
librosa.display.waveplot(data, sr=sampling_rate)
```

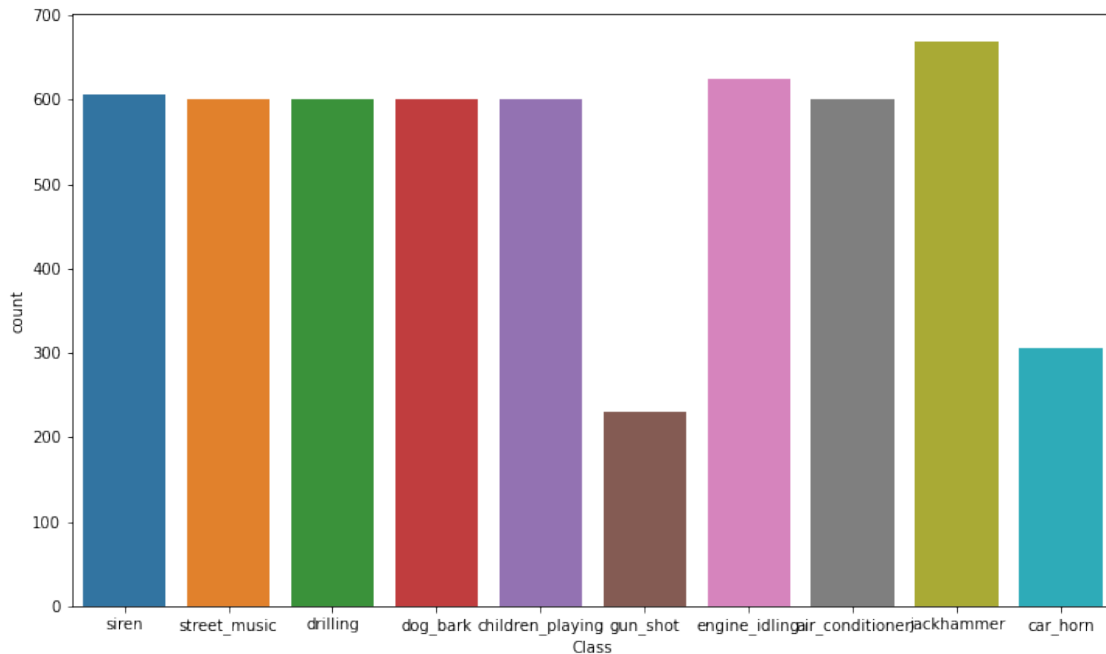
Class: car_horn

[30]: <matplotlib.collections.PolyCollection at 0x7f1354eb2438>



```
[34]: import seaborn as sns
plt.figure(figsize=(12,7))
sns.countplot(df['Class'])
```

[34]: <matplotlib.axes._subplots.AxesSubplot at 0x7f135451c278>



0.7 Input Split

```
[39]: import os

def parser(row):
    # path of the file
    file_name = os.path.join('Train', str(row.ID) + '.wav')
    # load the audio file
    x, sample_rate = librosa.load(file_name, res_type='kaiser_fast')
    # extract features from the data
    mfccs = np.mean(librosa.feature.mfcc(y=x, sr=sample_rate, n_mfcc=40).T,
    ↪axis=0)

    feature = mfccs
    label = row.Class

    return [feature, label]
```

```
[40]: data = df.apply(parser, axis=1)
      data.columns = ['feature', 'label']
```

```
[42]: data[0]
```

```
[42]: [array([-82.12358939, 139.50591598, -42.43086489, 24.82786139,
        -11.62076447, 23.49708426, -12.19458986, 25.89713885,
        -9.40527728, 21.21042898, -7.36882138, 14.25433903,
        -8.67870015, 7.75023765, -10.1241154 , 3.2581183 ,
        -11.35261914, 2.80096779, -7.04601346, 3.91331351,
        -2.3349743 , 2.01242254, -2.79394367, 4.12927394,
        -1.62076864, 4.32620082, -1.03440959, -1.23297714,
        -3.11085341, 0.32044827, -1.787786 , 0.44295495,
        -1.79164752, -0.76361758, -1.24246428, -0.27664012,
        0.65718559, -0.50237115, -2.60428533, -1.05346291]), 'siren']
```

```
[43]: # input split
      X = np.array(list(zip(*data))[0])
      y = np.array(list(zip(*data))[1])
```

0.8 Label encoder

```
[44]: from sklearn.preprocessing import LabelEncoder
      from keras.utils import np_utils

      le = LabelEncoder()
      y = np_utils.to_categorical(le.fit_transform(y))
```

```
[45]: y.shape
```

```
[45]: (5435, 10)
```

```
[46]: y[0]
```

```
[46]: array([0., 0., 0., 0., 0., 0., 0., 0., 1., 0.], dtype=float32)
```

0.9 Model Training

```
[54]: from keras.models import Sequential
      from keras.layers import Dense, Dropout, Activation, Flatten

      num_classes = 10

      # model creation
      model = Sequential()

      model.add(Dense(256, input_shape=(40,)))
```

```

model.add(Activation('relu'))
model.add(Dropout(0.3))

model.add(Dense(512))
model.add(Activation('relu'))
model.add(Dropout(0.3))

model.add(Dense(256))
model.add(Activation('relu'))
model.add(Dropout(0.3))

model.add(Dense(num_classes))
model.add(Activation('softmax'))

model.compile(loss='categorical_crossentropy', metrics='accuracy',
              optimizer='adam')

```

```

[55]: # train the model
model.fit(X, y, batch_size=32, epochs=100, validation_split=0.25)

```

```

Epoch 1/100
128/128 [=====] - 1s 6ms/step - loss: 10.9345 -
accuracy: 0.1495 - val_loss: 2.0777 - val_accuracy: 0.2597
Epoch 2/100
128/128 [=====] - 1s 5ms/step - loss: 2.1973 -
accuracy: 0.2605 - val_loss: 1.8058 - val_accuracy: 0.4091
Epoch 3/100
128/128 [=====] - 1s 5ms/step - loss: 1.9379 -
accuracy: 0.3309 - val_loss: 1.6051 - val_accuracy: 0.4643
Epoch 4/100
128/128 [=====] - 1s 5ms/step - loss: 1.7228 -
accuracy: 0.3807 - val_loss: 1.4608 - val_accuracy: 0.5077
Epoch 5/100
128/128 [=====] - 1s 5ms/step - loss: 1.5491 -
accuracy: 0.4630 - val_loss: 1.3355 - val_accuracy: 0.5453
Epoch 6/100
128/128 [=====] - 1s 5ms/step - loss: 1.4185 -
accuracy: 0.4843 - val_loss: 1.2097 - val_accuracy: 0.5938
Epoch 7/100
128/128 [=====] - 1s 5ms/step - loss: 1.3123 -
accuracy: 0.5384 - val_loss: 1.1305 - val_accuracy: 0.6350
Epoch 8/100
128/128 [=====] - 1s 5ms/step - loss: 1.2230 -
accuracy: 0.5772 - val_loss: 1.0233 - val_accuracy: 0.6843
Epoch 9/100
128/128 [=====] - 1s 5ms/step - loss: 1.0966 -
accuracy: 0.6136 - val_loss: 0.9303 - val_accuracy: 0.7145

```

Epoch 10/100
128/128 [=====] - 1s 5ms/step - loss: 1.0108 -
accuracy: 0.6537 - val_loss: 0.9144 - val_accuracy: 0.7020
Epoch 11/100
128/128 [=====] - 1s 6ms/step - loss: 0.9479 -
accuracy: 0.6794 - val_loss: 0.8196 - val_accuracy: 0.7513
Epoch 12/100
128/128 [=====] - 1s 5ms/step - loss: 0.8572 -
accuracy: 0.7047 - val_loss: 0.7841 - val_accuracy: 0.7623
Epoch 13/100
128/128 [=====] - 1s 5ms/step - loss: 0.7855 -
accuracy: 0.7370 - val_loss: 0.7249 - val_accuracy: 0.7815
Epoch 14/100
128/128 [=====] - 1s 5ms/step - loss: 0.7760 -
accuracy: 0.7312 - val_loss: 0.6986 - val_accuracy: 0.7881
Epoch 15/100
128/128 [=====] - 1s 5ms/step - loss: 0.7491 -
accuracy: 0.7465 - val_loss: 0.6526 - val_accuracy: 0.7969
Epoch 16/100
128/128 [=====] - 1s 5ms/step - loss: 0.7028 -
accuracy: 0.7673 - val_loss: 0.6324 - val_accuracy: 0.8035
Epoch 17/100
128/128 [=====] - 1s 5ms/step - loss: 0.6544 -
accuracy: 0.7700 - val_loss: 0.5943 - val_accuracy: 0.8197
Epoch 18/100
128/128 [=====] - 1s 5ms/step - loss: 0.6218 -
accuracy: 0.7817 - val_loss: 0.6158 - val_accuracy: 0.8116
Epoch 19/100
128/128 [=====] - 1s 5ms/step - loss: 0.5938 -
accuracy: 0.8014 - val_loss: 0.5872 - val_accuracy: 0.8094
Epoch 20/100
128/128 [=====] - 1s 5ms/step - loss: 0.5655 -
accuracy: 0.8001 - val_loss: 0.5640 - val_accuracy: 0.8197
Epoch 21/100
128/128 [=====] - 1s 5ms/step - loss: 0.5385 -
accuracy: 0.8160 - val_loss: 0.5352 - val_accuracy: 0.8322
Epoch 22/100
128/128 [=====] - 1s 5ms/step - loss: 0.5351 -
accuracy: 0.8151 - val_loss: 0.5039 - val_accuracy: 0.8411
Epoch 23/100
128/128 [=====] - 1s 5ms/step - loss: 0.5097 -
accuracy: 0.8191 - val_loss: 0.4905 - val_accuracy: 0.8506
Epoch 24/100
128/128 [=====] - 1s 5ms/step - loss: 0.4892 -
accuracy: 0.8348 - val_loss: 0.5233 - val_accuracy: 0.8521
Epoch 25/100
128/128 [=====] - 1s 5ms/step - loss: 0.5039 -
accuracy: 0.8224 - val_loss: 0.5153 - val_accuracy: 0.8484

Epoch 26/100
128/128 [=====] - 1s 5ms/step - loss: 0.4413 -
accuracy: 0.8531 - val_loss: 0.4805 - val_accuracy: 0.8514
Epoch 27/100
128/128 [=====] - 1s 5ms/step - loss: 0.4459 -
accuracy: 0.8451 - val_loss: 0.4908 - val_accuracy: 0.8550
Epoch 28/100
128/128 [=====] - 1s 5ms/step - loss: 0.4137 -
accuracy: 0.8596 - val_loss: 0.4724 - val_accuracy: 0.8528
Epoch 29/100
128/128 [=====] - 1s 5ms/step - loss: 0.4073 -
accuracy: 0.8548 - val_loss: 0.4560 - val_accuracy: 0.8793
Epoch 30/100
128/128 [=====] - 1s 5ms/step - loss: 0.4153 -
accuracy: 0.8538 - val_loss: 0.4550 - val_accuracy: 0.8653
Epoch 31/100
128/128 [=====] - 1s 5ms/step - loss: 0.4276 -
accuracy: 0.8559 - val_loss: 0.4321 - val_accuracy: 0.8631
Epoch 32/100
128/128 [=====] - 1s 5ms/step - loss: 0.3614 -
accuracy: 0.8824 - val_loss: 0.4599 - val_accuracy: 0.8779
Epoch 33/100
128/128 [=====] - 1s 5ms/step - loss: 0.3620 -
accuracy: 0.8733 - val_loss: 0.4363 - val_accuracy: 0.8720
Epoch 34/100
128/128 [=====] - 1s 5ms/step - loss: 0.3326 -
accuracy: 0.8882 - val_loss: 0.4184 - val_accuracy: 0.8779
Epoch 35/100
128/128 [=====] - 1s 5ms/step - loss: 0.3564 -
accuracy: 0.8765 - val_loss: 0.4316 - val_accuracy: 0.8823
Epoch 36/100
128/128 [=====] - 1s 5ms/step - loss: 0.3115 -
accuracy: 0.8869 - val_loss: 0.4323 - val_accuracy: 0.8801
Epoch 37/100
128/128 [=====] - 1s 5ms/step - loss: 0.3554 -
accuracy: 0.8836 - val_loss: 0.4180 - val_accuracy: 0.8764
Epoch 38/100
128/128 [=====] - 1s 5ms/step - loss: 0.3401 -
accuracy: 0.8867 - val_loss: 0.3973 - val_accuracy: 0.8904
Epoch 39/100
128/128 [=====] - 1s 5ms/step - loss: 0.3368 -
accuracy: 0.8922 - val_loss: 0.4650 - val_accuracy: 0.8631
Epoch 40/100
128/128 [=====] - 1s 5ms/step - loss: 0.3444 -
accuracy: 0.8800 - val_loss: 0.4376 - val_accuracy: 0.8896
Epoch 41/100
128/128 [=====] - 1s 5ms/step - loss: 0.2883 -
accuracy: 0.9044 - val_loss: 0.3772 - val_accuracy: 0.8889

Epoch 42/100
128/128 [=====] - 1s 5ms/step - loss: 0.3086 -
accuracy: 0.8952 - val_loss: 0.3987 - val_accuracy: 0.8904
Epoch 43/100
128/128 [=====] - 1s 5ms/step - loss: 0.2890 -
accuracy: 0.8943 - val_loss: 0.3914 - val_accuracy: 0.8859
Epoch 44/100
128/128 [=====] - 1s 5ms/step - loss: 0.2837 -
accuracy: 0.9025 - val_loss: 0.4280 - val_accuracy: 0.8911
Epoch 45/100
128/128 [=====] - 1s 5ms/step - loss: 0.3258 -
accuracy: 0.9006 - val_loss: 0.3884 - val_accuracy: 0.8852
Epoch 46/100
128/128 [=====] - 1s 6ms/step - loss: 0.2781 -
accuracy: 0.9124 - val_loss: 0.4229 - val_accuracy: 0.8940
Epoch 47/100
128/128 [=====] - 1s 5ms/step - loss: 0.3023 -
accuracy: 0.8968 - val_loss: 0.3803 - val_accuracy: 0.8926
Epoch 48/100
128/128 [=====] - 1s 5ms/step - loss: 0.2292 -
accuracy: 0.9182 - val_loss: 0.3708 - val_accuracy: 0.8999
Epoch 49/100
128/128 [=====] - 1s 5ms/step - loss: 0.2638 -
accuracy: 0.9092 - val_loss: 0.3868 - val_accuracy: 0.8874
Epoch 50/100
128/128 [=====] - 1s 5ms/step - loss: 0.2714 -
accuracy: 0.9109 - val_loss: 0.4067 - val_accuracy: 0.8874
Epoch 51/100
128/128 [=====] - 1s 5ms/step - loss: 0.2402 -
accuracy: 0.9187 - val_loss: 0.3848 - val_accuracy: 0.8992
Epoch 52/100
128/128 [=====] - 1s 5ms/step - loss: 0.2717 -
accuracy: 0.9139 - val_loss: 0.3934 - val_accuracy: 0.8926
Epoch 53/100
128/128 [=====] - 1s 5ms/step - loss: 0.2731 -
accuracy: 0.9102 - val_loss: 0.3777 - val_accuracy: 0.9036
Epoch 54/100
128/128 [=====] - 1s 5ms/step - loss: 0.2456 -
accuracy: 0.9216 - val_loss: 0.4055 - val_accuracy: 0.8845
Epoch 55/100
128/128 [=====] - 1s 5ms/step - loss: 0.2561 -
accuracy: 0.9116 - val_loss: 0.3938 - val_accuracy: 0.9043
Epoch 56/100
128/128 [=====] - 1s 5ms/step - loss: 0.2459 -
accuracy: 0.9220 - val_loss: 0.3679 - val_accuracy: 0.8999
Epoch 57/100
128/128 [=====] - 1s 5ms/step - loss: 0.2536 -
accuracy: 0.9210 - val_loss: 0.3828 - val_accuracy: 0.8955

Epoch 58/100
128/128 [=====] - 1s 5ms/step - loss: 0.2276 -
accuracy: 0.9189 - val_loss: 0.3920 - val_accuracy: 0.9021
Epoch 59/100
128/128 [=====] - 1s 5ms/step - loss: 0.2219 -
accuracy: 0.9285 - val_loss: 0.3921 - val_accuracy: 0.9051
Epoch 60/100
128/128 [=====] - 1s 5ms/step - loss: 0.2325 -
accuracy: 0.9197 - val_loss: 0.4338 - val_accuracy: 0.9021
Epoch 61/100
128/128 [=====] - 1s 5ms/step - loss: 0.2177 -
accuracy: 0.9242 - val_loss: 0.3954 - val_accuracy: 0.8999
Epoch 62/100
128/128 [=====] - 1s 5ms/step - loss: 0.2323 -
accuracy: 0.9280 - val_loss: 0.4031 - val_accuracy: 0.8933
Epoch 63/100
128/128 [=====] - 1s 5ms/step - loss: 0.2495 -
accuracy: 0.9152 - val_loss: 0.4132 - val_accuracy: 0.9058
Epoch 64/100
128/128 [=====] - 1s 5ms/step - loss: 0.2013 -
accuracy: 0.9313 - val_loss: 0.4326 - val_accuracy: 0.8940
Epoch 65/100
128/128 [=====] - 1s 5ms/step - loss: 0.2480 -
accuracy: 0.9251 - val_loss: 0.3966 - val_accuracy: 0.9088
Epoch 66/100
128/128 [=====] - 1s 5ms/step - loss: 0.2965 -
accuracy: 0.9232 - val_loss: 0.3706 - val_accuracy: 0.9065
Epoch 67/100
128/128 [=====] - 1s 5ms/step - loss: 0.2011 -
accuracy: 0.9338 - val_loss: 0.3669 - val_accuracy: 0.9102
Epoch 68/100
128/128 [=====] - 1s 5ms/step - loss: 0.1873 -
accuracy: 0.9384 - val_loss: 0.4369 - val_accuracy: 0.8962
Epoch 69/100
128/128 [=====] - 1s 5ms/step - loss: 0.2712 -
accuracy: 0.9157 - val_loss: 0.3485 - val_accuracy: 0.9080
Epoch 70/100
128/128 [=====] - 1s 5ms/step - loss: 0.2294 -
accuracy: 0.9224 - val_loss: 0.3572 - val_accuracy: 0.9102
Epoch 71/100
128/128 [=====] - 1s 5ms/step - loss: 0.1905 -
accuracy: 0.9372 - val_loss: 0.3791 - val_accuracy: 0.9029
Epoch 72/100
128/128 [=====] - 1s 5ms/step - loss: 0.1879 -
accuracy: 0.9304 - val_loss: 0.3752 - val_accuracy: 0.9095
Epoch 73/100
128/128 [=====] - 1s 5ms/step - loss: 0.2025 -
accuracy: 0.9359 - val_loss: 0.4290 - val_accuracy: 0.8970

Epoch 74/100
128/128 [=====] - 1s 5ms/step - loss: 0.1840 -
accuracy: 0.9373 - val_loss: 0.3987 - val_accuracy: 0.9014
Epoch 75/100
128/128 [=====] - 1s 5ms/step - loss: 0.2182 -
accuracy: 0.9357 - val_loss: 0.4522 - val_accuracy: 0.8948
Epoch 76/100
128/128 [=====] - 1s 5ms/step - loss: 0.2166 -
accuracy: 0.9372 - val_loss: 0.3313 - val_accuracy: 0.9183
Epoch 77/100
128/128 [=====] - 1s 5ms/step - loss: 0.2009 -
accuracy: 0.9352 - val_loss: 0.4012 - val_accuracy: 0.9051
Epoch 78/100
128/128 [=====] - 1s 5ms/step - loss: 0.1828 -
accuracy: 0.9363 - val_loss: 0.3910 - val_accuracy: 0.9095
Epoch 79/100
128/128 [=====] - 1s 5ms/step - loss: 0.1728 -
accuracy: 0.9475 - val_loss: 0.3741 - val_accuracy: 0.9051
Epoch 80/100
128/128 [=====] - 1s 6ms/step - loss: 0.1792 -
accuracy: 0.9395 - val_loss: 0.3704 - val_accuracy: 0.9102
Epoch 81/100
128/128 [=====] - 1s 5ms/step - loss: 0.1769 -
accuracy: 0.9407 - val_loss: 0.3389 - val_accuracy: 0.9191
Epoch 82/100
128/128 [=====] - 1s 5ms/step - loss: 0.2019 -
accuracy: 0.9305 - val_loss: 0.4008 - val_accuracy: 0.9021
Epoch 83/100
128/128 [=====] - 1s 5ms/step - loss: 0.1888 -
accuracy: 0.9357 - val_loss: 0.3751 - val_accuracy: 0.9146
Epoch 84/100
128/128 [=====] - 1s 5ms/step - loss: 0.2088 -
accuracy: 0.9357 - val_loss: 0.3838 - val_accuracy: 0.9021
Epoch 85/100
128/128 [=====] - 1s 5ms/step - loss: 0.1858 -
accuracy: 0.9421 - val_loss: 0.3945 - val_accuracy: 0.9110
Epoch 86/100
128/128 [=====] - 1s 5ms/step - loss: 0.2087 -
accuracy: 0.9345 - val_loss: 0.4125 - val_accuracy: 0.9058
Epoch 87/100
128/128 [=====] - 1s 5ms/step - loss: 0.1704 -
accuracy: 0.9406 - val_loss: 0.3750 - val_accuracy: 0.9169
Epoch 88/100
128/128 [=====] - 1s 5ms/step - loss: 0.1565 -
accuracy: 0.9485 - val_loss: 0.3929 - val_accuracy: 0.9095
Epoch 89/100
128/128 [=====] - 1s 5ms/step - loss: 0.1564 -
accuracy: 0.9501 - val_loss: 0.3995 - val_accuracy: 0.9058

```

Epoch 90/100
128/128 [=====] - 1s 5ms/step - loss: 0.1976 -
accuracy: 0.9330 - val_loss: 0.3827 - val_accuracy: 0.9139
Epoch 91/100
128/128 [=====] - 1s 5ms/step - loss: 0.1643 -
accuracy: 0.9475 - val_loss: 0.3867 - val_accuracy: 0.9095
Epoch 92/100
128/128 [=====] - 1s 5ms/step - loss: 0.2080 -
accuracy: 0.9333 - val_loss: 0.3833 - val_accuracy: 0.9051
Epoch 93/100
128/128 [=====] - 1s 5ms/step - loss: 0.1938 -
accuracy: 0.9378 - val_loss: 0.3767 - val_accuracy: 0.9073
Epoch 94/100
128/128 [=====] - 1s 5ms/step - loss: 0.2045 -
accuracy: 0.9322 - val_loss: 0.3608 - val_accuracy: 0.9191
Epoch 95/100
128/128 [=====] - 1s 5ms/step - loss: 0.1565 -
accuracy: 0.9552 - val_loss: 0.3429 - val_accuracy: 0.9117
Epoch 96/100
128/128 [=====] - 1s 5ms/step - loss: 0.1859 -
accuracy: 0.9393 - val_loss: 0.3583 - val_accuracy: 0.9213
Epoch 97/100
128/128 [=====] - 1s 5ms/step - loss: 0.1961 -
accuracy: 0.9441 - val_loss: 0.3974 - val_accuracy: 0.9088
Epoch 98/100
128/128 [=====] - 1s 5ms/step - loss: 0.1798 -
accuracy: 0.9384 - val_loss: 0.4386 - val_accuracy: 0.8992
Epoch 99/100
128/128 [=====] - 1s 5ms/step - loss: 0.2149 -
accuracy: 0.9316 - val_loss: 0.4006 - val_accuracy: 0.9110
Epoch 100/100
128/128 [=====] - 1s 5ms/step - loss: 0.1530 -
accuracy: 0.9526 - val_loss: 0.3809 - val_accuracy: 0.9169

```

```
[55]: <tensorflow.python.keras.callbacks.History at 0x7f12c2d88cc0>
```

```
[ ]:
```