

# Data Collection

## Download the Dataset

# Image Pre-Processing

## Importing the Necessary Libraries

In [1]:

```
pwd
```

Out[1]:

```
 '/home/wsuser/work'
```

In [2]:

```
!pip install imutils
```

```
Requirement already satisfied: imutils in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (0.5.4)
```

In [3]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import zipfile as zf
import os
import random
import cv2
import pickle
from imutils import build_montages
from imutils import paths
from sklearn.metrics import classification_report, confusion_matrix
from sklearn import metrics
from sklearn.preprocessing import LabelEncoder, LabelBinarizer
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier, ExtraTreesClassifier
from skimage import feature
```

In [4]:

```
sns.set()
os.getcwd()
```

Out[4]:

```
 '/home/wsuser/work'
```

## Loading the training and testing dataset

In [5]:

```
import os, types
import pandas as pd
from botocore.client import Config
import ibm_boto3

def __iter__(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
cos_client = ibm_boto3.client(service_name='s3',
                              ibm_api_key_id='9PBhSes0z9VA9j6pKAN6AEHF8eukFhE19WfNRaxepkC5',
                              ibm_auth_endpoint="https://iam.cloud.ibm.com/oidc/token",
                              config=Config(signature_version='oauth'),
                              endpoint_url='https://s3.private.us.cloud-object-storage.appdomain.cloud')

bucket = 'parkinsonsprediction-donotdelete-pr-tpmhf0fv6vnyvw'
object_key = 'spiral-20221112T063807Z-001.zip'

streaming_body_5 = cos_client.get_object(Bucket=bucket, Key=object_key)['Body']

# Your data file was loaded into a botocore.response.StreamingBody object.
# Please read the documentation of ibm_boto3 and pandas to learn more about the possibilities to load the data.
# ibm_boto3 documentation: https://ibm.github.io/ibm-cos-sdk-python/
# pandas documentation: http://pandas.pydata.org/
```

In [6]:

```
from io import BytesIO
import zipfile
unzip = zipfile.ZipFile(BytesIO(streaming_body_5.read()), 'r')
file_paths = unzip.namelist()
```

```
for path in file_paths:
    unzip.extract(path)
```

In [7]: pwd

Out[7]: '/home/wsuser/work'

In [8]: filenames = os.listdir('/home/wsuser/work/spiral/training/')

In [9]: spiral\_train\_healthy = os.listdir('/home/wsuser/work/spiral/training/healthy/')  
spiral\_train\_park = os.listdir('/home/wsuser/work/spiral/training/parkinson/')  
  
fp\_spiral\_train\_healthy = '/home/wsuser/work/spiral/training/healthy/'  
fp\_spiral\_train\_park = '/home/wsuser/work/spiral/training/parkinson/'  
  
spiral\_test\_healthy = os.listdir('/home/wsuser/work/spiral/testing/healthy/')  
spiral\_test\_park = os.listdir('/home/wsuser/work/spiral/testing/parkinson/')  
  
fp\_spiral\_test\_healthy = '/home/wsuser/work/spiral/testing/healthy/'  
fp\_spiral\_test\_park = '/home/wsuser/work/spiral/testing/parkinson/'

## Quantifying Images

In [10]: def quantify\_image(image):  
features = feature.hog(image,orientations=9,  
pixels\_per\_cell=(10,10),cells\_per\_block=(2,2),transform\_sqrt=True,block\_norm="L1")  
  
return features

## Splitting up of training and testing data

In [11]: trainX = []  
testX = []  
outputs = []  
trainY = []  
testY = []  
  
for i in spiral\_train\_healthy:  
image = cv2.imread(fp\_spiral\_train\_healthy+i)  
image = cv2.cvtColor(image , cv2.COLOR\_BGR2GRAY)  
image = cv2.resize(image , (200,200))  
image =cv2.threshold(image, 0, 255,cv2.THRESH\_BINARY\_INV | cv2.THRESH\_OTSU)[1]  
features = quantify\_image(image)  
trainX.append(features)  
trainY.append('healthy')  
  
for i in spiral\_train\_park:  
image = cv2.imread(fp\_spiral\_train\_park+i)  
image = cv2.cvtColor(image , cv2.COLOR\_BGR2GRAY)  
image = cv2.resize(image , (200,200))  
image = cv2.threshold(image ,0,255,cv2.THRESH\_BINARY\_INV | cv2.THRESH\_OTSU)[1]  
features = quantify\_image(image)  
trainX.append(features)  
trainY.append('parkinson')  
  
for i in spiral\_test\_healthy:  
image = cv2.imread(fp\_spiral\_test\_healthy+i)  
outputs.append(image)  
image = cv2.cvtColor(image , cv2.COLOR\_BGR2GRAY)  
image = cv2.resize(image , (200,200))  
image = cv2.threshold(image ,0,255,cv2.THRESH\_BINARY\_INV | cv2.THRESH\_OTSU)[1]  
features = quantify\_image(image)  
testX.append(features)  
testY.append('healthy')  
  
for i in spiral\_test\_park:  
image = cv2.imread(fp\_spiral\_test\_park+i)  
outputs.append(image)  
image = cv2.cvtColor(image , cv2.COLOR\_BGR2GRAY)  
image = cv2.resize(image , (200,200))  
image = cv2.threshold(image ,0,255,cv2.THRESH\_BINARY\_INV | cv2.THRESH\_OTSU)[1]  
features = quantify\_image(image)  
testX.append(features)  
testY.append('parkinson')

```
In [12]: trainX = np.array(trainX)
          testX = np.array(testX)
          trainY = np.array(trainY)
          testY = np.array(testY)
          trainX
```

```
Out[12]: array([[0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.],
               ...,
               [0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.]])
```

```
In [13]: trainY
```

[illegible]

```
In [14]: testX
```

```
Out[14]: array([[0., 0., 0., ..., 0., 0., 0.],
                [0., 0., 0., ..., 0., 0., 0.],
                [0., 0., 0., ..., 0., 0., 0.],
                ...,
                [0., 0., 0., ..., 0., 0., 0.],
                [0., 0., 0., ..., 0., 0., 0.],
                [0., 0., 0., ..., 0., 0., 0.]])
```

```
In [15]: testY
```

```
Out[15]: array(['healthy', 'healthy', 'healthy', 'healthy', 'healthy', 'healthy',
                'healthy', 'healthy', 'healthy', 'healthy', 'healthy', 'healthy',
                'healthy', 'healthy', 'healthy', 'parkinson', 'parkinson',
                'parkinson', 'parkinson', 'parkinson', 'parkinson', 'parkinson',
                'parkinson', 'parkinson', 'parkinson', 'parkinson', 'parkinson',
                'parkinson', 'parkinson', 'parkinson'], dtype='<U9')
```

## Label Encoding

```
In [16]: le = LabelEncoder()
trainY = le.fit_transform(trainY)
testY = le.transform(testY)
print(trainX.shape, trainY.shape)

(72, 12996) (72,)
```

```
In [17]: trainY
```

```
Out[17]: array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
                0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1,  
                1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,  
                1, 1, 1, 1, 1, 1])
```

```
In [18]: testY
```

```
Out[18]: array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1,
                1, 1, 1, 1, 1, 1, 1, 1, 1])
```

## Model Building

## Training the model

```
In [38]: print("Training model....")
          model = RandomForestClassifier(n_estimators=100)
          model.fit(trainX,trainY)
```

Training model....

```
Out[38]: RandomForestClassifier()
```

```
In [39]: preds = model.predict(testX)
preds
```

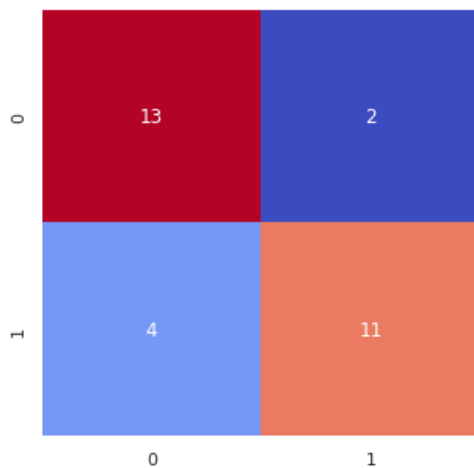
```
Out[39]: array([1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1,
        1, 0, 1, 1, 1, 0, 0, 1])
```

## Model Evaluation

```
In [40]: cnf = confusion_matrix(testY,preds)
cnf
```

```
Out[40]: array([[13,  2],
        [ 4, 11]])
```

```
In [33]: plt.figure(figsize=(5,5))
sns.heatmap(cnf , annot=True , cmap="coolwarm" , cbar=False)
plt.show()
```



```
In [34]: acc = metrics.accuracy_score(testY,preds)
acc
```

```
Out[34]: 0.8
```

```
In [35]: indexes = np.random.randint(0,30,25)
indexes
```

```
Out[35]: array([ 5,  4,  9, 14, 25, 19,  5, 25,  1, 22, 10, 26,  1, 13, 17, 24, 28,
        13, 20, 12,  9, 11, 27, 10, 22])
```

## Testing the Model

```
In [36]: testpath=list(paths.list_images(fp_spiral_train_healthy))
idxs=np.arange(0,len(testpath))
idxs=np.random.choice(idxs,size=(25,),replace=False)
images=[]

for i in idxs:
    image=cv2.imread(testpath[i])
    output=image.copy()
    output=cv2.resize(output,(128,128))
    image=cv2.cvtColor(image,cv2.COLOR_BGR2GRAY)
    image=cv2.resize(image,(200,200))
    image=cv2.threshold(image,0,255,cv2.THRESH_BINARY_INV | cv2.THRESH_OTSU)[1]

    features= quantify_image(image)
    preds=model.predict([features])
    label=le.inverse_transform(preds)[0]
    if label=="healthy":
        color=(0,255,0)
    else:
        color=(0,0,255)
    cv2.putText(output,label, (3,20),cv2.FONT_HERSHEY_SIMPLEX,0.5,color,2)
    images.append(output)
```

```
In [95]: !pip install opencv-python
```

```
|██████████████████████| 60.9 MB 25.9 MB/s eta 0:00:01
```

```
Requirement already satisfied: numpy>=1.14.5 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from opencv-python) (1.20.3)
```

```
Installing collected packages: opencv-python
```

Successfully installed opencv-python-4.6.0.66

```
In [96]: montage = build_montages(images, (128,128), (5,5))[0]
cv2.imshow("Output",montage)
cv2.waitKey(0)
'''
```

```
montage=build_montages(images,(128,128),(5,5))[0]
cv2_imshow(montage)
cv2.waitKey(0)
'''
```

[illegible]

```
/tmp/wsuser/ipykernel_164/3795390558.py in <module>
```

```
1 montage = build_montages(images,(128,128),(5,5))[0]
```

```
----> 2 cv2.imshow("Output",montage)
```

```
3 cv2.waitKey(0)
```

4 ' ' '

```
5 montage=build_montages(images,(128,128),(5,5))[0]
```

```
error: OpenCV(4.5.5) ../modules/highgui/src/window.cpp:1268: error: (-2:Unspecified error) The function is not implemented. Rebuild the library with Windows, GTK+ 2.x or Cocoa support. If you are on Ubuntu or Debian, install libgtk2.0-dev and pkg-config, then re-run cmake or configure script in function 'cvShowImage'
```

## Predicting the model-Accuracy and Confusion Matrix

```
In [37]: predictions = model.predict(testX)

cm = confusion_matrix(testY, predictions).flatten()
print(cm)
(tn, fp, fn, tp) = cm
accuracy = (tp + tn) / float(cm.sum())
print(accuracy)
```

```
[13  2  4 11]
0.8
```

## Save the Model

```
In [41]: pickle.dump(model, open('parkinson.pkl', 'wb'))
```

```
In [42]: !tar -zcvf parkinsons-model_new.tgz parkinson.pkl
```

parkinson.pkl

```
In [43]: ls
```

parkinson.pkl parkinsons-model\_new.tgz parkinsons\_model.tar.gz [spiral/](#)

```
In [44]: !pip install watson-machine-learning-client --upgrade
```

Requirement already satisfied: watson-machine-learning-client in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (1.0.391)

Requirement already satisfied: requests in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from watson-machine-learning-client) (2.26.0)

Requirement already satisfied: pandas in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from watson-machine-learning-client) (1.3.4)

Requirement already satisfied: certifi in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from watson-machine-learning-client) (2022.9.24)

Requirement already satisfied: urllib3 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from watson-machine-learning-client) (1.26.7)

Requirement already satisfied: tqdm in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from watson-machine-learning-client) (4.62.3)

Requirement already satisfied: tabulate in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from watson-machine-learning-client) (0.8.9)

Requirement already satisfied: boto3 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from watson-machine-learning-client) (1.18.21)

Requirement already satisfied: ibm-cos-sdk in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from watson-machine-learning-client) (2.11.0)

Requirement already satisfied: lomond in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from watson-machine-learning-client) (0.3.3)

Requirement already satisfied: botocore<1.22.0,>=1.21.21 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from boto3->watson-machine-learning-client) (1.21.41)

Requirement already satisfied: s3transfer<0.6.0,>=0.5.0 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from boto3->watson-machine-learning-client) (0.5.0)

Requirement already satisfied: jmespath<1.0.0,>=0.7.1 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from boto3->watson-machine-learning-client) (0.10.0)

Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from botocore<1.22.0,>=1.21.21->boto3->watson-machine-learning-client) (2.8.2)

Requirement already satisfied: six>=1.5 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from python-dateutil<3.0.0,>=2.1->botocore<1.22.0,>=1.21.21->boto3->watson-machine-learning-client) (1.15.0)

Requirement already satisfied: ibm-cos-sdk-s3transfer==2.11.0 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from ibm-cos-sdk->watson-machine-learning-client) (2.11.0)

Requirement already satisfied: ibm-cos-sdk-core==2.11.0 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from ibm-cos-sdk->watson-machine-learning-client) (2.11.0)

Requirement already satisfied: idna<4,>=2.5 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from requests->watson-machine-learning-client) (3.3)

Requirement already satisfied: charset-normalizer~=2.0.0 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from requests->watson-machine-learning-client) (2.0.4)

Requirement already satisfied: pytz>=2017.3 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from pandas->watson-machine-learning-client) (2021.3)

Requirement already satisfied: numpy>=1.17.3 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from pandas->watson-machine-learning-client) (1.20.3)

```
In [45]: from ibm_watson_machine_learning import APIClient
wml_credentials = {
    "url": "https://us-south.ml.cloud.ibm.com",
    "apikey": "qCp1VQAqMTPikiYRALscUVshlebbUbGLitQkAiWaBtIz"
}
client = APIClient(wml_credentials)
```

```
In [46]: def guid_from_space_name(client, space_name):
        space = client.spaces.get_details()
        return(next(item for item in space['resources'] if item['entity']['name']==space_name)['metadata']['id'])
```

```
In [47]: space_uid = guid_from_space_name(client, 'ParkinsonsPrediction_Space')
print(space_uid)

1c89246d-042b-40ae-a78b-1f8446569702
```

```
In [48]: client.set.default_space(space_uid)
```

```
Out[48]: 'SUCCESS'
```

```
In [49]: client.software_specifications.list()
```

NAME	ASSET_ID	TYPE
default_py3.6	0062b8c9-8b7d-44a0-a9b9-46c416adcbd9	base
kernel-spark3.2-scala2.12	020d69ce-7ac1-5e68-ac1a-31189867356a	base
pytorch-onnx_1.3-py3.7-edt	069ea134-3346-5748-b513-49120e15d288	base
scikit-learn_0.20-py3.6	09c5a1d0-9c1e-4473-a344-eb7b665ff687	base
spark-mllib_3.0-scala_2.12	09f4cff0-90a7-5899-b9ed-1ef348aebdee	base
pytorch-onnx_rt22.1-py3.9	0b848dd4-e681-5599-be41-b5f6fccc6471	base
ai-function_0.1-py3.6	0cdb0f1e-5376-4f4d-92dd-da3b69aa9bda	base
shiny-r3.6	0e6e79df-875e-4f24-8ae9-62dcc2148306	base
tensorflow_2.4-py3.7-horovod	1092590a-307d-563d-9b62-4eb7d64b3f22	base
pytorch_1.1-py3.6	10ac12d6-6b30-4ccd-8392-3e922c096a92	base
tensorflow_1.15-py3.6-ddl	111e41b3-de2d-5422-a4d6-bf776828c4b7	base
autoai-kb_rt22.2-py3.10	125b6d9a-5b1f-5e8d-972a-b251688ccf40	base
runtime-22.1-py3.9	12b83a17-24d8-5082-900f-0ab31fbfd3cb	base
scikit-learn_0.22-py3.6	154010fa-5b3b-4ac1-82af-4d5ee5abbc85	base
default_r3.6	1b70aec3-ab34-4b87-8aa0-a4a3c8296a36	base
pytorch-onnx_1.3-py3.6	1bc6029a-cc97-56da-b8e0-39c3880dbbe7	base
kernel-spark3.3-r3.6	1c9e5454-f216-59dd-a20e-474a5cdf5988	base
pytorch-onnx_rt22.1-py3.9-edt	1d362186-7ad5-5b59-8b6c-9d0880bde37f	base
tensorflow_2.1-py3.6	1eb25b84-d6ed-5dde-b6a5-3fbdf1665666	base
spark-mllib_3.2	20047f72-0a98-58c7-9ff5-a77b012eb8f5	base
tensorflow_2.4-py3.8-horovod	217c16f6-178f-56bf-824a-b19f20564c49	base
runtime-22.1-py3.9-cuda	26215f05-08c3-54a1-a1b0-da66306ce658	base
do_py3.8	295addb5-9ef9-547e-9bf4-92ae3563e720	base
autoai-ts_3.8-py3.8	2aa0c932-798f-5ae9-abd6-15e0c2402fb5	base
tensorflow_1.15-py3.6	2b73a275-7cbf-420b-a912-eae7f436e0bc	base
kernel-spark3.3-py3.9	2b7961e2-e3b1-5a8c-a491-482c8368839a	base
pytorch_1.2-py3.6	2c8ef57d-2687-4b7d-acce-01f94976dac1	base
spark-mllib_2.3	2e51f700-bca0-4b0d-88dc-5c6791338875	base
pytorch-onnx_1.1-py3.6-edt	32983cea-3f32-4400-8965-dde874a8d67e	base
spark-mllib_3.0-py37	36507ebe-8770-55ba-ab2a-eafe787600e9	base
spark-mllib_2.4	390d21f8-e58b-4fac-9c55-d7ceda621326	base
autoai-ts_rt22.2-py3.10	396b2e83-0953-5b86-9a55-7ce1628a406f	base
xgboost_0.82-py3.6	39e31acd-5f30-41dc-ae44-60233c80306e	base
pytorch-onnx_1.2-py3.6-edt	40589d0e-7019-4e28-8daa-fb03b6f4fe12	base
pytorch-onnx_rt22.2-py3.10	40e73f55-783a-5535-b3fa-0c8b94291431	base
default_r36py38	41c247d3-45f8-5a71-b065-8580229facf0	base
autoai-ts_rt22.1-py3.9	4269d26e-07ba-5d40-8f66-2d495b0c71f7	base
autoai-obm_3.0	42b92e18-d9ab-567f-988a-4240ba1ed5f7	base
pmm1-3.0_4.3	493bcb95-16f1-5bc5-bee8-81b8af80e9c7	base
spark-mllib_2.4-r_3.6	49403dff-92e9-4c87-a3d7-a42d0021c095	base
xgboost_0.90-py3.6	4ff8d6c2-1343-4c18-85e1-689c965304d3	base
pytorch-onnx_1.1-py3.6	50f95b2a-bc16-43bb-bc94-b0bed208c60b	base
autoai-ts_3.9-py3.8	52c57136-80fa-572e-8728-a5e7cbb42cde	base
spark-mllib_2.4-scala_2.11	55a70f99-7320-4be5-9fb9-9edb5a443af5	base
spark-mllib_3.0	5c1b0ca2-4977-5c2e-9439-ffd44ea8ffe9	base
autoai-obm_2.0	5c2e37fa-80b8-5e77-840f-d912469614ee	base
spss-modeler_18.1	5c3cad7e-507f-4b2a-a9a3-ab53a21dee8b	base
cuda-py3.8	5d3232bf-c86b-5df4-a2cd-7bb870a1cd4e	base
autoai-kb_3.1-py3.7	632d4b22-10aa-5180-88f0-f52dfb6444d7	base
pytorch-onnx_1.7-py3.8	634d3cdc-b562-5bf9-a2d4-ea90a478456b	base

Note: Only first 50 records were displayed. To display more use 'limit' parameter.

```
In [50]: software_spec_uid = client.software_specifications.get_uid_by_name('runtime-22.1-py3.9')
software_spec_uid
```

```
Out[50]: '12b83a17-24d8-5082-900f-0ab31fbfd3cb'
```

```
In [51]: model_details = client.repository.store_model(model='parkinsons-model_new.tgz', meta_props={
    client.repository.ModelMetaNames.NAME: "RandomForest",
    client.repository.ModelMetaNames.TYPE: "xgboost_1.5",
    client.repository.ModelMetaNames.SOFTWARE_SPEC_UID: software_spec_uid
})
model_id = client.repository.get_model_uid(model_details)
```

This method is deprecated, please use get\_model\_id()

/opt/conda/envs/Python-3.9/lib/python3.9/site-packages/ibm\_watson\_machine\_learning/repository.py:1453: UserWarning:  
This method is deprecated, please use get\_model\_id()  
warn("This method is deprecated, please use get\_model\_id()")

```
In [53]: model_id
```

```
Out[53]: '552baa6f-c154-4ddd-b38b-43db30b63439'
```

```
In [54]: client.repository.download("552baa6f-c154-4ddd-b38b-43db30b63439", 'parkinsons.tar.gz')
```

Successfully saved model content to file: 'parkinsons.tar.gz'

```
Out[54]: '/home/wsuser/work/parkinsons.tar.gz'
```