In [50]:
```
import pandas as pd
data=pd.read_csv('D:\\download\\blended_effnets.csv')
```

In [51]:
```
data
```

Out[51]:

	image_name	target
0	ISIC_0052060	0.027242
1	ISIC_0052349	0.027575
2	ISIC_0058510	0.027162
3	ISIC_0073313	0.028097
4	ISIC_0073502	0.036457
...	...	...
10977	ISIC_9992485	0.025966
10978	ISIC_9996992	0.038696
10979	ISIC_9997917	0.086362
10980	ISIC_9998234	0.028173
10981	ISIC_9999302	0.071839

10982 rows × 2 columns

In [52]:
```
#to get the information about the structure of dataset
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10982 entries, 0 to 10981
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   image_name  10982 non-null  object 
 1   target      10982 non-null  float64
dtypes: float64(1), object(1)
memory usage: 171.7+ KB
```

In [53]:
```
data.shape
```

Out[53]:
```
(10982, 2)
```

In [54]:
```
data.size
```

Out[54]:
```
21964
```

In [55]:
```
data.ndim
```

Out[55]:
```
2
```

In [56]:
```
#Exploratory data Analysis:
#Checking Structure of data
#missing values
#data transformation
#statistical analysis
```

In [57]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10982 entries, 0 to 10981
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   image_name  10982 non-null  object 
 1   target      10982 non-null  float64
dtypes: float64(1), object(1)
memory usage: 171.7+ KB
```

In [58]:
```
#check datatypes inconsistency
data.dtypes
```

Out[58]:
```
image_name     object
target        float64
dtype: object
```

In [59]:
```
#checking the data type with missing values
#NA
#NAN
#NULL
#isna()
#isnull()
```

In [60]:
```
#cheking missing value
data.isna().sum()
```

Out[60]:
```
image_name    0
target        0
dtype: int64
```

In [61]:
```
data.dropna()
```

Out[61]:

	image_name	target
0	ISIC_0052060	0.027242
1	ISIC_0052349	0.027575
2	ISIC_0058510	0.027162
3	ISIC_0073313	0.028097
4	ISIC_0073502	0.036457
...	...	...
10977	ISIC_9992485	0.025966
10978	ISIC_9996992	0.038696
10979	ISIC_9997917	0.086362
10980	ISIC_9998234	0.028173
10981	ISIC_9999302	0.071839

10982 rows × 2 columns

In [62]:
```
data.dropna(axis=0,inplace=True)
```

In [63]:
```
data.isna().sum()
```

Out[63]:
```
image_name    0
target        0
dtype: int64
```

In [64]:
```
data
```

Out[64]:

	image_name	target
0	ISIC_0052060	0.027242
1	ISIC_0052349	0.027575
2	ISIC_0058510	0.027162
3	ISIC_0073313	0.028097
4	ISIC_0073502	0.036457
...	...	...
10977	ISIC_9992485	0.025966
10978	ISIC_9996992	0.038696
10979	ISIC_9997917	0.086362
10980	ISIC_9998234	0.028173
10981	ISIC_9999302	0.071839

10982 rows × 2 columns

In [65]:
```
data.fillna(0,inplace=True)
```

In [66]:
```
data.isna().sum()
```

Out[66]:
```
image_name    0
target        0
dtype: int64
```

In [67]:
```
#extract first 50 records
data.head()
```

Out[67]:

	image_name	target
0	ISIC_0052060	0.027242
1	ISIC_0052349	0.027575
2	ISIC_0058510	0.027162
3	ISIC_0073313	0.028097
4	ISIC_0073502	0.036457

In [68]:
```
data.head(50)
```

Out[68]:

	image_name	target
0	ISIC_0052060	0.027242
1	ISIC_0052349	0.027575
2	ISIC_0058510	0.027162
3	ISIC_0073313	0.028097
4	ISIC_0073502	0.036457
5	ISIC_0074618	0.028636
6	ISIC_0076801	0.029146
7	ISIC_0077586	0.032823
8	ISIC_0082004	0.031738
9	ISIC_0082785	0.033315
10	ISIC_0085332	0.024194
11	ISIC_0087243	0.024236
12	ISIC_0088167	0.027416
13	ISIC_0089356	0.026044
14	ISIC_0090283	0.038399
15	ISIC_0092481	0.031827
16	ISIC_0095455	0.028145
17	ISIC_0095790	0.025301
18	ISIC_0097719	0.059419
19	ISIC_0098348	0.055910
20	ISIC_0101040	0.029956
21	ISIC_0101383	0.025984
22	ISIC_0105104	0.130169
23	ISIC_0108196	0.023657
24	ISIC_0112420	0.506590
25	ISIC_0114232	0.025025
26	ISIC_0115798	0.031286
27	ISIC_0116844	0.027284
28	ISIC_0126432	0.027598
29	ISIC_0131485	0.026011
30	ISIC_0131912	0.026673
31	ISIC_0131978	0.025874
32	ISIC_0139322	0.028724
33	ISIC_0142066	0.035664
34	ISIC_0143109	0.029742
35	ISIC_0148210	0.037295
36	ISIC_0148937	0.030558
37	ISIC_0150213	0.026533
38	ISIC_0152553	0.025738
39	ISIC_0153127	0.032131
40	ISIC_0154299	0.030035
41	ISIC_0155813	0.069606
42	ISIC_0155983	0.242346
43	ISIC_0157088	0.031336
44	ISIC_0158761	0.026901
45	ISIC_0161963	0.038800
46	ISIC_0165178	0.040000
47	ISIC_0165230	0.033253
48	ISIC_0165615	0.026227
49	ISIC_0165617	0.027598

In [69]:
```
data.tail()
```

Out[69]:

	image_name	target
10977	ISIC_9992485	0.025966
10978	ISIC_9996992	0.038696
10979	ISIC_9997917	0.086362
10980	ISIC_9998234	0.028173
10981	ISIC_9999302	0.071839

In [70]:
```
data.columns
```

Out[70]:
```
Index(['image_name', 'target'], dtype='object')
```

In [71]:
```
#number of uniquevalues present for the dataset
data.nunique()
```

Out[71]:
```
image_name    10982
target        10980
dtype: int64
```

In [72]:
```
#to delete a record we use drop() function
data.columns
```

Out[72]:
```
Index(['image_name', 'target'], dtype='object')
```

In [73]:
```
data.drop(['image_name'],axis=1,inplace=True)
```

In [74]:
```
data.columns
```

Out[74]:
```
Index(['target'], dtype='object')
```

In [75]:
```
#to delete with a given range of rows
data.drop(data.index[10:20],inplace=True)
```

In [77]:
```
#statistical operations on dataset
data
```

Out[77]:

	target
0	0.027242
1	0.027575
2	0.027162
3	0.028097
4	0.036457
...	...
10977	0.025966
10978	0.038696
10979	0.086362
10980	0.028173
10981	0.071839

10972 rows × 1 columns

In [78]:
```
data.dtypes
```

Out[78]:
```
target    float64
dtype: object
```

In [ ]:
```

```