logist_regg

March 13, 2024

```
[1]: import pandas as pd
     import numpy as np
```

```
[2]: cancer_data=pd.read_csv('/content/cancer.csv')
```

```
[3]: cancer_data.head()
```

```
[3]:    index        id  radius_mean  texture_mean  perimeter_mean  area_mean  \
     0      1    842302        17.99         10.38          122.80     1001.0
     1      2    842517        20.57         17.77          132.90     1326.0
     2      3  84300903        19.69         21.25          130.00     1203.0
     3      4  84348301        11.42         20.38           77.58      386.1
     4      5  84358402        20.29         14.34          135.10     1297.0

        smoothness_mean  compactness_mean  concavity_mean  concave_points_mean  \
     0          0.11840           0.27760          0.3001              0.14710
     1          0.08474           0.07864          0.0869              0.07017
     2          0.10960           0.15990          0.1974              0.12790
     3          0.14250           0.28390          0.2414              0.10520
     4          0.10030           0.13280          0.1980              0.10430

        …  smoothness_worst  compactness_worst  concavity_worst  \
     0  …            0.1622             0.6656           0.7119
     1  …            0.1238             0.1866           0.2416
     2  …            0.1444             0.4245           0.4504
     3  …            0.2098             0.8663           0.6869
     4  …            0.1374             0.2050           0.4000

        concave points_worst  symmetry_worst  fractal_dimension_worst  N Stage  \
     0                0.2654          0.4601                  0.11890       N1
     1                0.1860          0.2750                  0.08902       N2
     2                0.2430          0.3613                  0.08758       N3
     3                0.2575          0.6638                  0.17300       N1
     4                0.1625          0.2364                  0.07678       N1

        6th Stage              differentiate  diagnosis
     0        IIA      Poorly differentiated          M
```

```
1       IIIA  Moderately differentiated          M
2       IIIC  Moderately differentiated          M
3        IIA      Poorly differentiated          M
4        IIB      Poorly differentiated          M

[5 rows x 36 columns]
```

[4]:
```
cancer_data.shape
```

[4]:
```
(569, 36)
```

[5]:
```
cancer_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 36 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   index                    569 non-null    int64
 1   id                       569 non-null    int64
 2   radius_mean              569 non-null    float64
 3   texture_mean             569 non-null    float64
 4   perimeter_mean           569 non-null    float64
 5   area_mean                569 non-null    float64
 6   smoothness_mean          569 non-null    float64
 7   compactness_mean         569 non-null    float64
 8   concavity_mean           552 non-null    float64
 9   concave_points_mean      550 non-null    float64
 10  symmetry_mean            568 non-null    float64
 11  fractal_dimension        569 non-null    float64
 12  radius_se                569 non-null    float64
 13  texture_se               569 non-null    float64
 14  perimeter_se             569 non-null    float64
 15  area_se                  569 non-null    float64
 16  smoothness_se            569 non-null    float64
 17  compactness_se           569 non-null    float64
 18  concavity_se             554 non-null    float64
 19  concave_points_se        550 non-null    float64
 20  symmetry_se              569 non-null    float64
 21  fractal_dimension_se     569 non-null    float64
 22  radius_worst             569 non-null    float64
 23  texture_worst            569 non-null    float64
 24  texture_worst.1          569 non-null    float64
 25  area_worst               569 non-null    float64
 26  smoothness_worst         569 non-null    float64
 27  compactness_worst        568 non-null    float64
 28  concavity_worst          562 non-null    float64
 29  concave points_worst     563 non-null    float64
 30  symmetry_worst           569 non-null    float64
 31  fractal_dimension_worst  569 non-null    float64
 32  N Stage                  569 non-null    object
 33  6th Stage                569 non-null    object
 34  differentiate            569 non-null    object
 35  diagnosis                569 non-null    object
dtypes: float64(30), int64(2), object(4)
memory usage: 160.2+ KB
```

[6]:
```
cancer_data.tail()
```

[6]:
```
     index      id  radius_mean  texture_mean  perimeter_mean  area_mean  \
564    565  926424        21.56         22.39          142.00     1479.0
565    566  926682        20.13         28.25          131.20     1261.0
566    567  926954        16.60         28.08          108.30      858.1
567    568  927241        20.60         29.33          140.10     1265.0
568    569   92751         7.76         24.54           47.92      181.0

     smoothness_mean  compactness_mean  concavity_mean  concave_points_mean  \
564          0.11100           0.11590         0.24390              0.13890
565          0.09780           0.10340         0.14400              0.09791
566          0.08455           0.10230         0.09251              0.05302
567          0.11780           0.27700         0.35140              0.15200
568          0.05263           0.04362             NaN                  NaN

     …  smoothness_worst  compactness_worst  concavity_worst  \
564  …           0.14100            0.21130           0.4107
565  …           0.11660            0.19220           0.3215
566  …           0.11390            0.30940           0.3403
567  …           0.16500            0.86810           0.9387
568  …           0.08996            0.06444              NaN

     concave points_worst  symmetry_worst  fractal_dimension_worst  N Stage  \
564                0.2216          0.2060                  0.07115       N2
565                0.1628          0.2572                  0.06637       N2
566                0.1418          0.2218                  0.07820       N1
567                0.2650          0.4087                  0.12400       N1
568                   NaN          0.2871                  0.07039       N3

     6th Stage              differentiate  diagnosis
564       IIIA  Moderately differentiated          M
565       IIIA      Poorly differentiated          M
566        IIA        Well differentiated          M
567        IIA        Well differentiated          M
568       IIIC        Well differentiated          B
```

```
[5 rows x 36 columns]
```

[7]:

```
cancer_data.isnull().sum()
```

[7]:

```
index                       0
id                          0
radius_mean                 0
texture_mean                0
perimeter_mean              0
area_mean                   0
smoothness_mean             0
compactness_mean            0
concavity_mean             17
concave_points_mean        19
symmetry_mean               1
fractal_dimension           0
radius_se                   0
texture_se                  0
perimeter_se                0
area_se                     0
smoothness_se               0
compactness_se              0
concavity_se               15
concave_points_se          19
symmetry_se                 0
fractal_dimension_se        0
radius_worst                0
texture_worst               0
texture_worst.1             0
area_worst                  0
smoothness_worst            0
compactness_worst           1
concavity_worst             7
concave points_worst        6
symmetry_worst              0
fractal_dimension_worst     0
N Stage                     0
6th Stage                   0
differentiate               0
diagnosis                   0
dtype: int64
```

[8]:

```
cancer_data['concavity_mean'].fillna(cancer_data['concavity_mean'].
  ↪median(),inplace=True)
```

[9]:

```
cancer_data['concave_points_se'].fillna(cancer_data['concave_points_se'].
  ↪median(),inplace=True)
```

```
cancer_data['concavity_mean'].fillna(cancer_data['concavity_mean'].
 ↪median(),inplace=True)
cancer_data['concave_points_mean'].fillna(cancer_data['concave_points_mean'].
 ↪median(),inplace=True)
```

[10]:
```
cancer_data['concavity_worst'].fillna(cancer_data['concavity_worst'].mode().
 ↪values[0],inplace=True)
# if we have column with string datatype then replace null values with␣
 ↪mode,since mode represent object so do value[0]
```

[11]:
```
cancer_data['concave points_worst'].fillna(cancer_data['concave points_worst'].
 ↪mode().values[0],inplace=True)
cancer_data['compactness_worst'].fillna(cancer_data['compactness_worst'].mode().
 ↪values[0],inplace=True)
```

[12]:
```
cancer_data.isnull().sum()
```

[12]:
```
index                     0
id                        0
radius_mean               0
texture_mean              0
perimeter_mean            0
area_mean                 0
smoothness_mean           0
compactness_mean          0
concavity_mean            0
concave_points_mean       0
symmetry_mean             1
fractal_dimension         0
radius_se                 0
texture_se                0
perimeter_se              0
area_se                   0
smoothness_se             0
compactness_se            0
concavity_se             15
concave_points_se         0
symmetry_se               0
fractal_dimension_se      0
radius_worst              0
texture_worst             0
texture_worst.1           0
area_worst                0
smoothness_worst          0
compactness_worst         0
concavity_worst           0
concave points_worst      0
```

```
symmetry_worst              0
fractal_dimension_worst     0
N Stage                     0
6th Stage                   0
differentiate               0
diagnosis                   0
dtype: int64
```

[13]:

```
cancer_data['symmetry_mean'].fillna(cancer_data['symmetry_mean'].
 ↪median(),inplace=True)
cancer_data['concavity_se'].fillna(cancer_data['concavity_se'].
 ↪median(),inplace=True)
```

[14]:

```
cancer_data.isnull().sum()
```

[14]:

```
index                      0
id                         0
radius_mean                0
texture_mean               0
perimeter_mean             0
area_mean                  0
smoothness_mean            0
compactness_mean           0
concavity_mean             0
concave_points_mean        0
symmetry_mean              0
fractal_dimension          0
radius_se                  0
texture_se                 0
perimeter_se               0
area_se                    0
smoothness_se              0
compactness_se             0
concavity_se               0
concave_points_se          0
symmetry_se                0
fractal_dimension_se       0
radius_worst               0
texture_worst              0
texture_worst.1            0
area_worst                 0
smoothness_worst           0
compactness_worst          0
concavity_worst            0
concave points_worst       0
symmetry_worst             0
fractal_dimension_worst    0
```

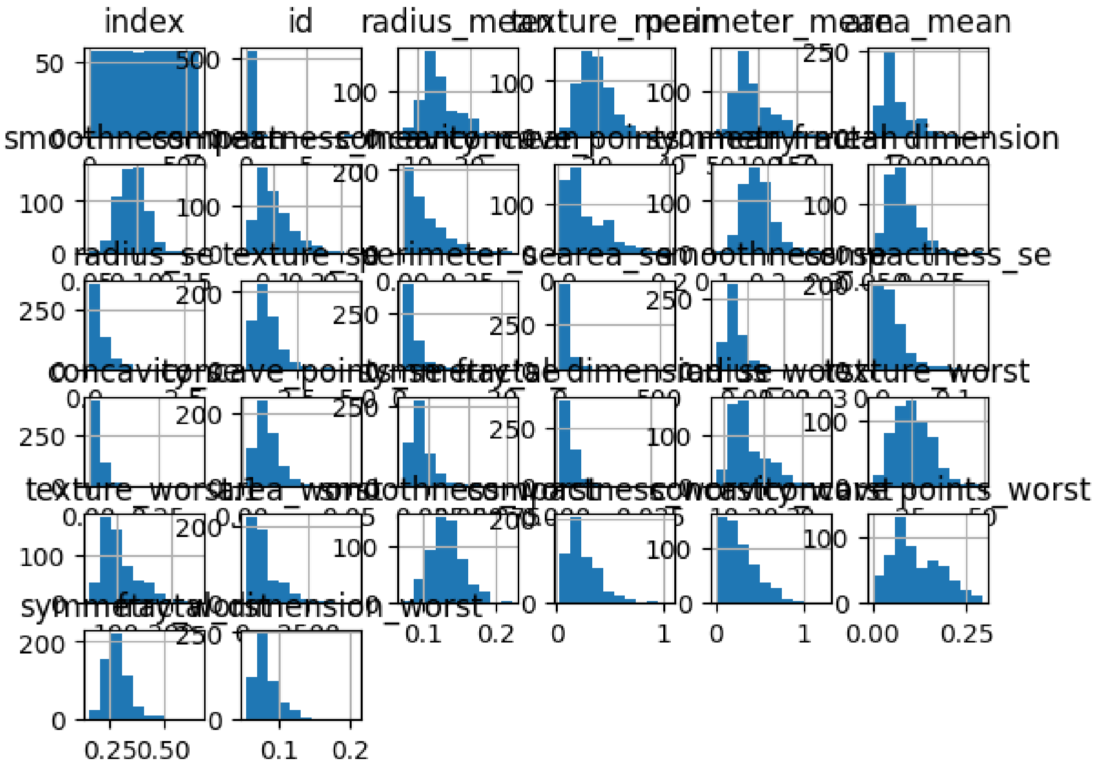
```
N Stage                    0
6th Stage                  0
differentiate              0
diagnosis                  0
dtype: int64
```

[15]:

```
# now complete the non-normal distribution data into normal-distribution data
cancer_data.hist()
```

[15]:

```
array([[<Axes: title={'center': 'index'}>,
        <Axes: title={'center': 'id'}>,
        <Axes: title={'center': 'radius_mean'}>,
        <Axes: title={'center': 'texture_mean'}>,
        <Axes: title={'center': 'perimeter_mean'}>,
        <Axes: title={'center': 'area_mean'}>],
       [<Axes: title={'center': 'smoothness_mean'}>,
        <Axes: title={'center': 'compactness_mean'}>,
        <Axes: title={'center': 'concavity_mean'}>,
        <Axes: title={'center': 'concave_points_mean'}>,
        <Axes: title={'center': 'symmetry_mean'}>,
        <Axes: title={'center': 'fractal_dimension'}>],
       [<Axes: title={'center': 'radius_se'}>,
        <Axes: title={'center': 'texture_se'}>,
        <Axes: title={'center': 'perimeter_se'}>,
        <Axes: title={'center': 'area_se'}>,
        <Axes: title={'center': 'smoothness_se'}>,
        <Axes: title={'center': 'compactness_se'}>],
       [<Axes: title={'center': 'concavity_se'}>,
        <Axes: title={'center': 'concave_points_se'}>,
        <Axes: title={'center': 'symmetry_se'}>,
        <Axes: title={'center': 'fractal_dimension_se'}>,
        <Axes: title={'center': 'radius_worst'}>,
        <Axes: title={'center': 'texture_worst'}>],
       [<Axes: title={'center': 'texture_worst.1'}>,
        <Axes: title={'center': 'area_worst'}>,
        <Axes: title={'center': 'smoothness_worst'}>,
        <Axes: title={'center': 'compactness_worst'}>,
        <Axes: title={'center': 'concavity_worst'}>,
        <Axes: title={'center': 'concave points_worst'}>],
       [<Axes: title={'center': 'symmetry_worst'}>,
        <Axes: title={'center': 'fractal_dimension_worst'}>, <Axes: >,
        <Axes: >, <Axes: >, <Axes: >]], dtype=object)
```



[16]:
```
cancer_data.skew()
```

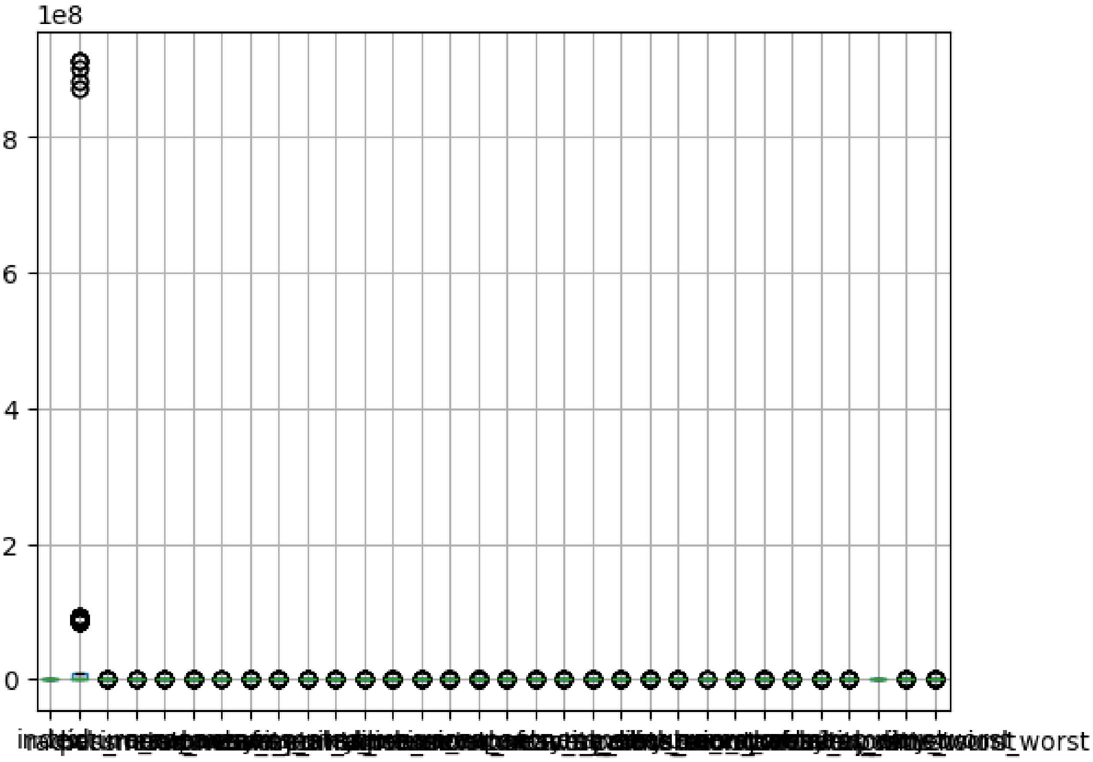
```
<ipython-input-16-a3ad186019e1>:1: FutureWarning: The default value of
numeric_only in DataFrame.skew is deprecated. In a future version, it will
default to False. In addition, specifying 'numeric_only=None' is deprecated.
Select only valid columns or specify the value of numeric_only to silence this
warning.
  cancer_data.skew()
```

[16]:
```
index                       0.000000
id                          6.473752
radius_mean                 0.942380
texture_mean                0.650450
perimeter_mean              0.990650
area_mean                   1.645732
smoothness_mean             0.456324
compactness_mean            1.190123
concavity_mean              1.449348
concave_points_mean         1.233690
symmetry_mean               0.725645
fractal_dimension           1.304489
radius_se                   3.088612
```

```
texture_se                 1.646444
perimeter_se               3.443615
area_se                    5.447186
smoothness_se              2.314450
compactness_se             1.902221
concavity_se               5.285191
concave_points_se          1.708892
symmetry_se                2.195133
fractal_dimension_se       3.923969
radius_worst               1.103115
texture_worst              0.498321
texture_worst.1            1.128164
area_worst                 1.859373
smoothness_worst           0.415426
compactness_worst          1.479888
concavity_worst            1.149806
concave points_worst       0.492616
symmetry_worst             1.433928
fractal_dimension_worst    1.662579
dtype: float64
```

[17]:
```
cancer_data.boxplot()
```

[17]:
```
<Axes: >
```



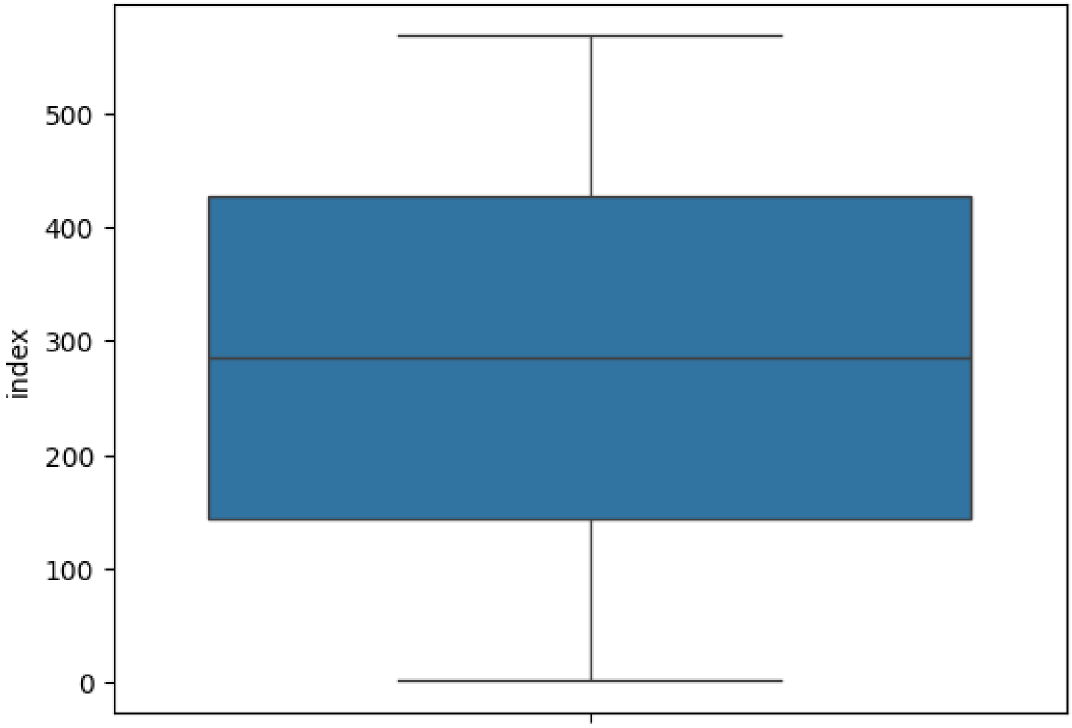
[18]:

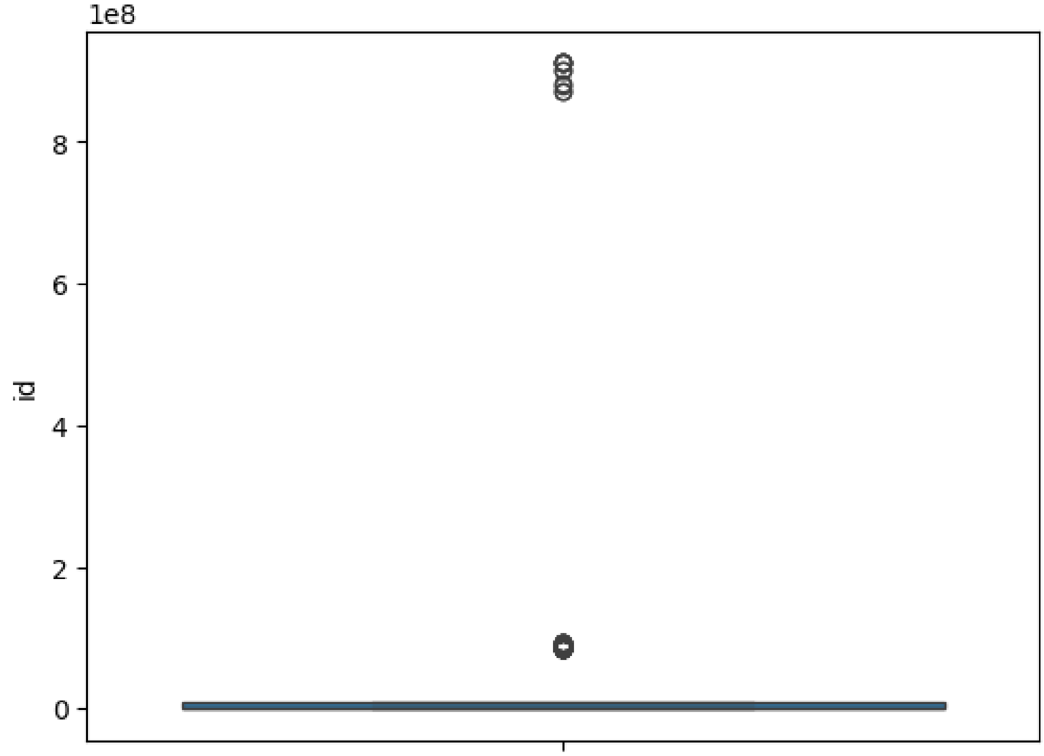
```
# removing outliers
import seaborn as sns
import matplotlib.pyplot as plt
```

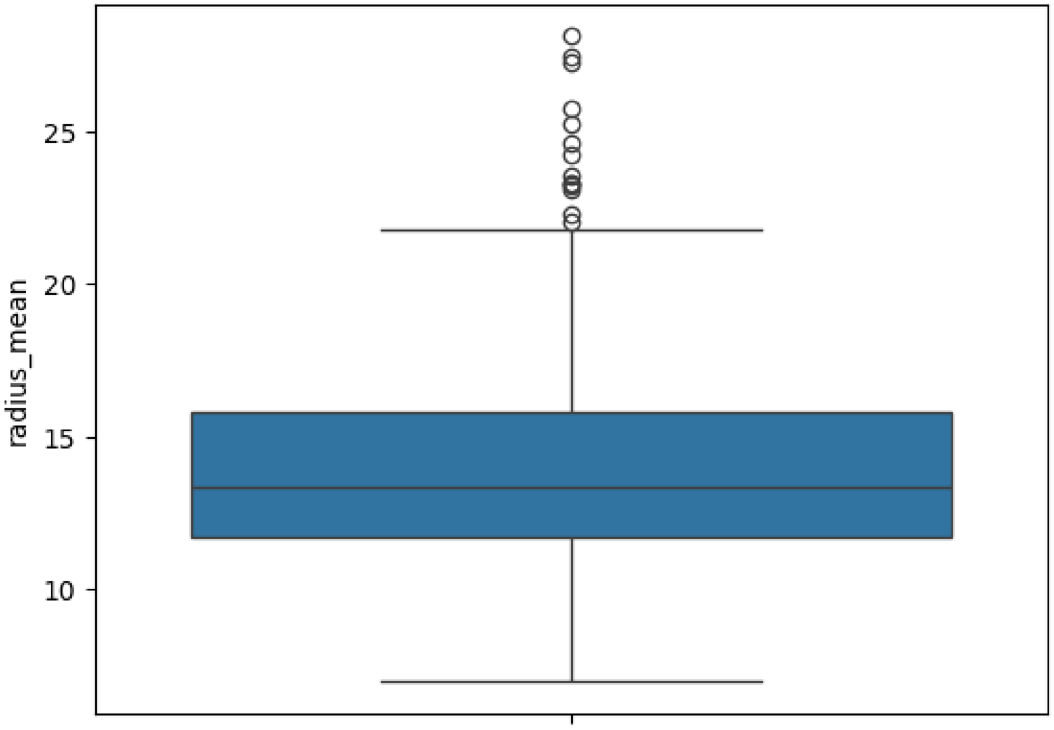
[19]:

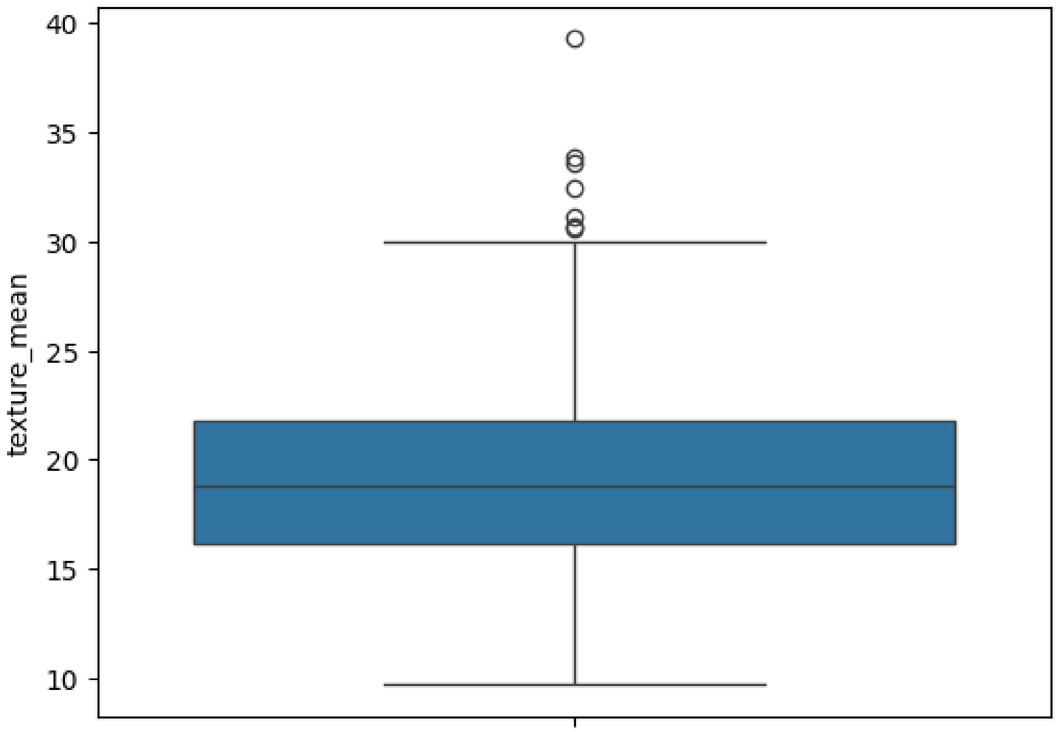
```
numerical_cols=cancer_data.select_dtypes(include='number').columns
for col in numerical_cols:
  sns.boxplot(cancer_data[col])

  # show plot
  plt.show()
```

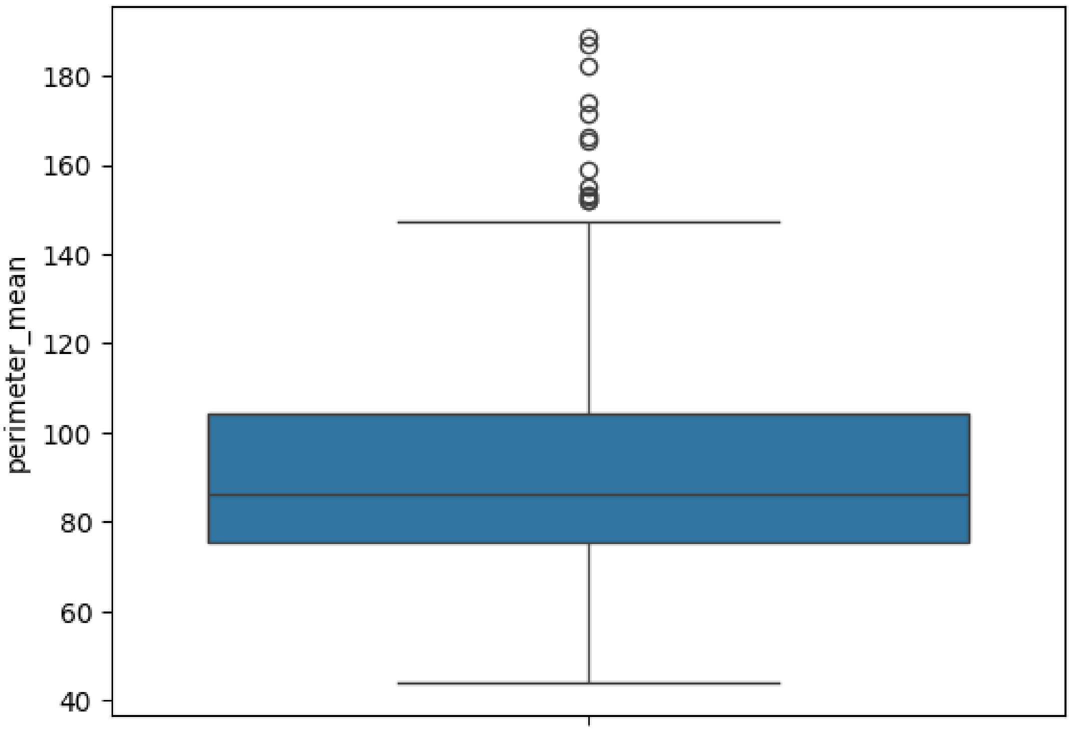



1e8
8
6
4
2
0
id

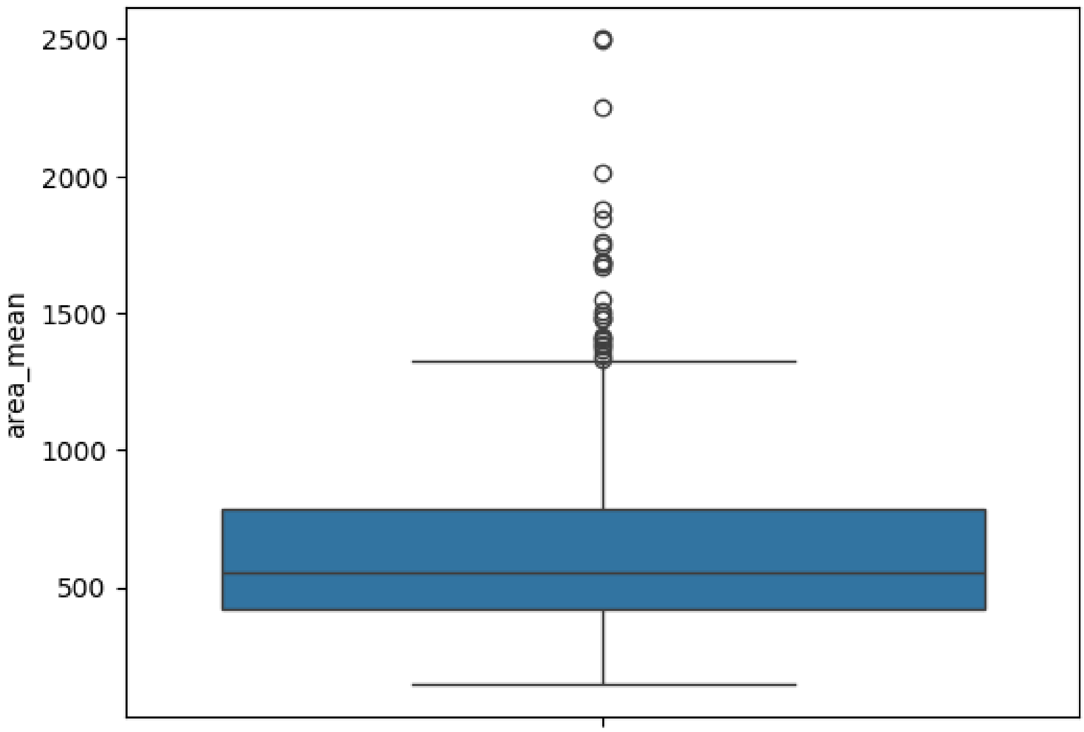

25
20
15
10
radius_mean

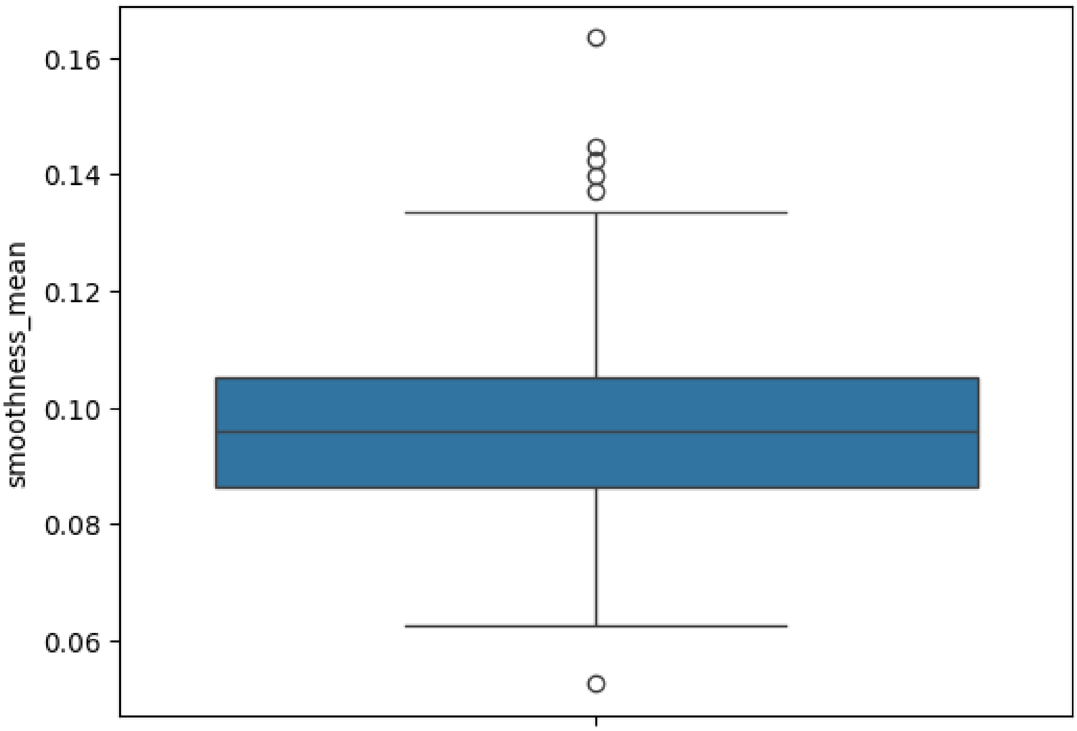


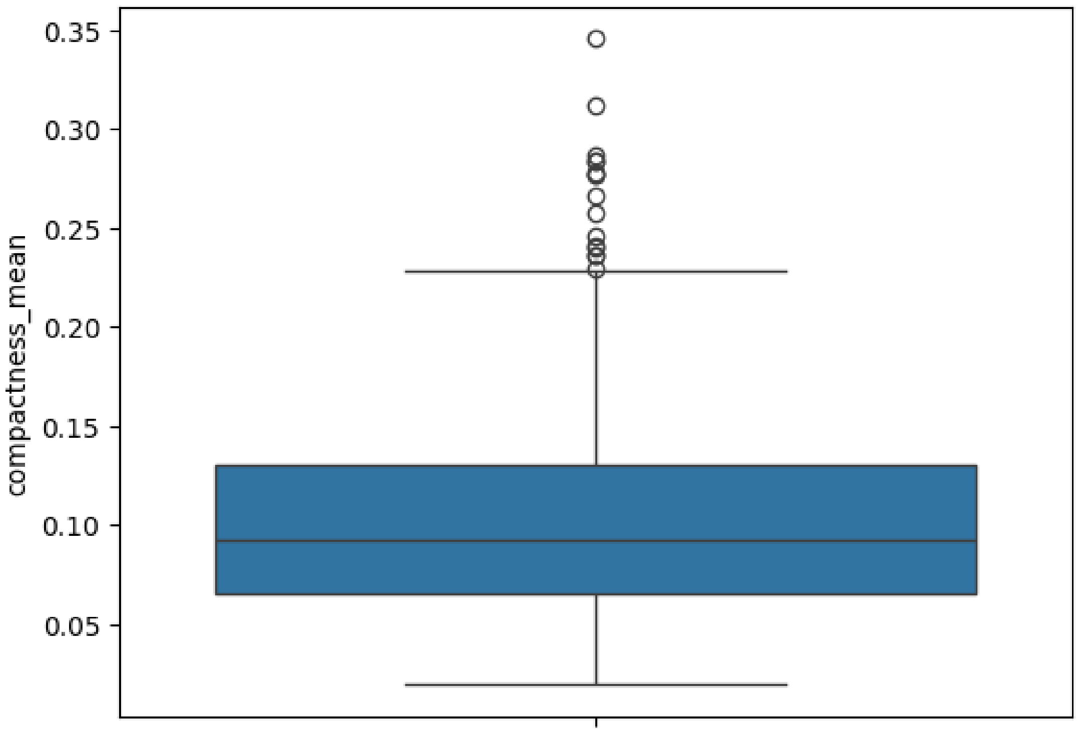
40
35
30
25
20
15
10
texture_mean

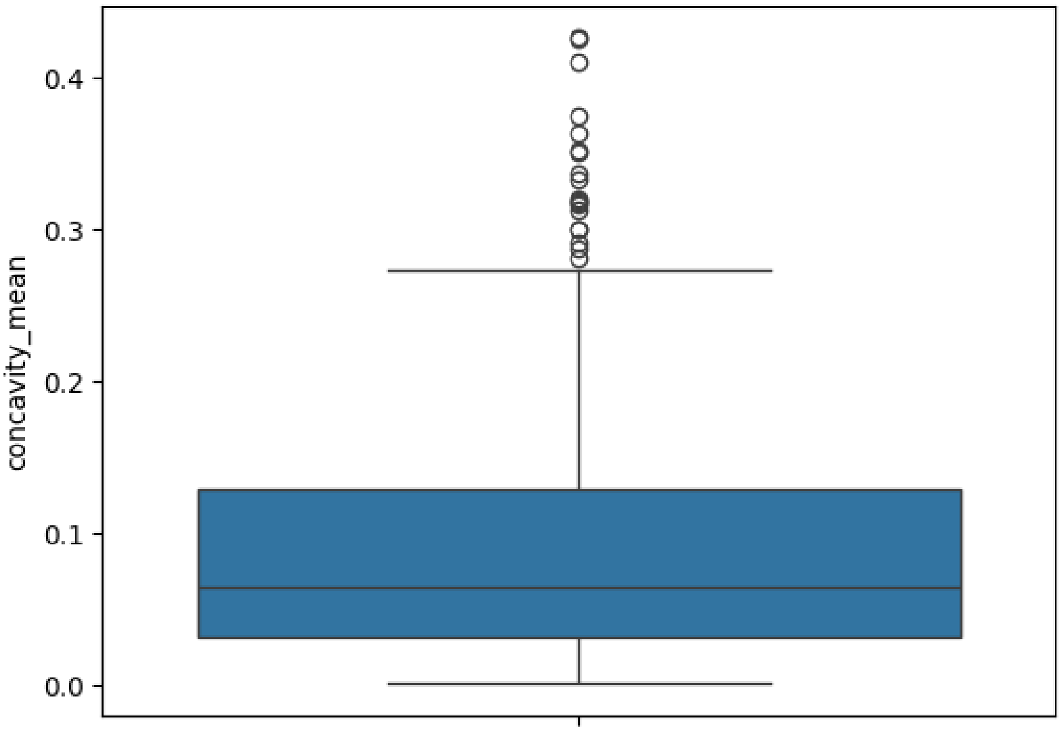


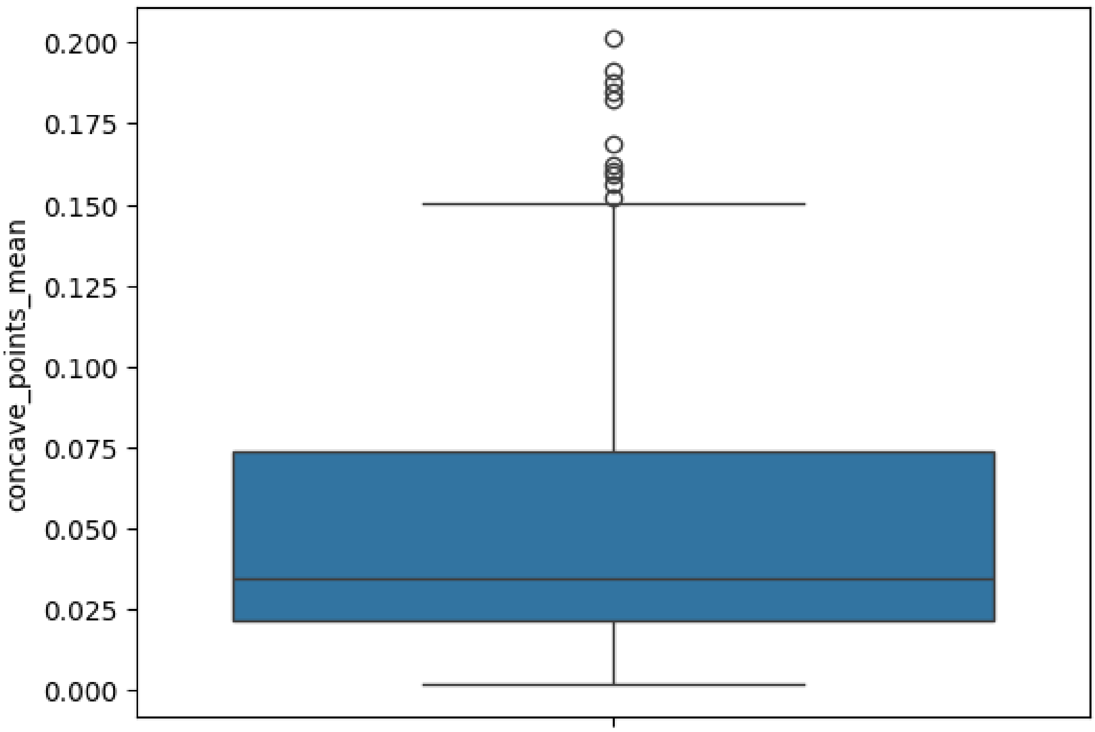
180
160
140
120
100
80
60
40
perimeter_mean

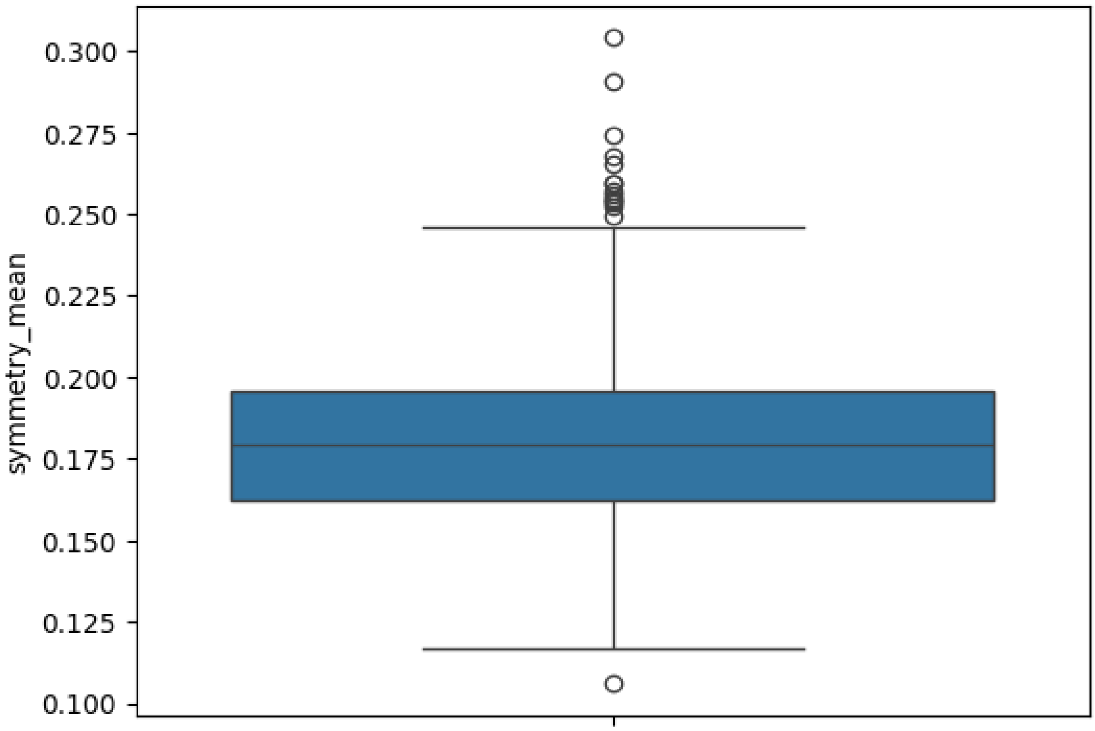

2500
2000
1500
1000
500
area_mean

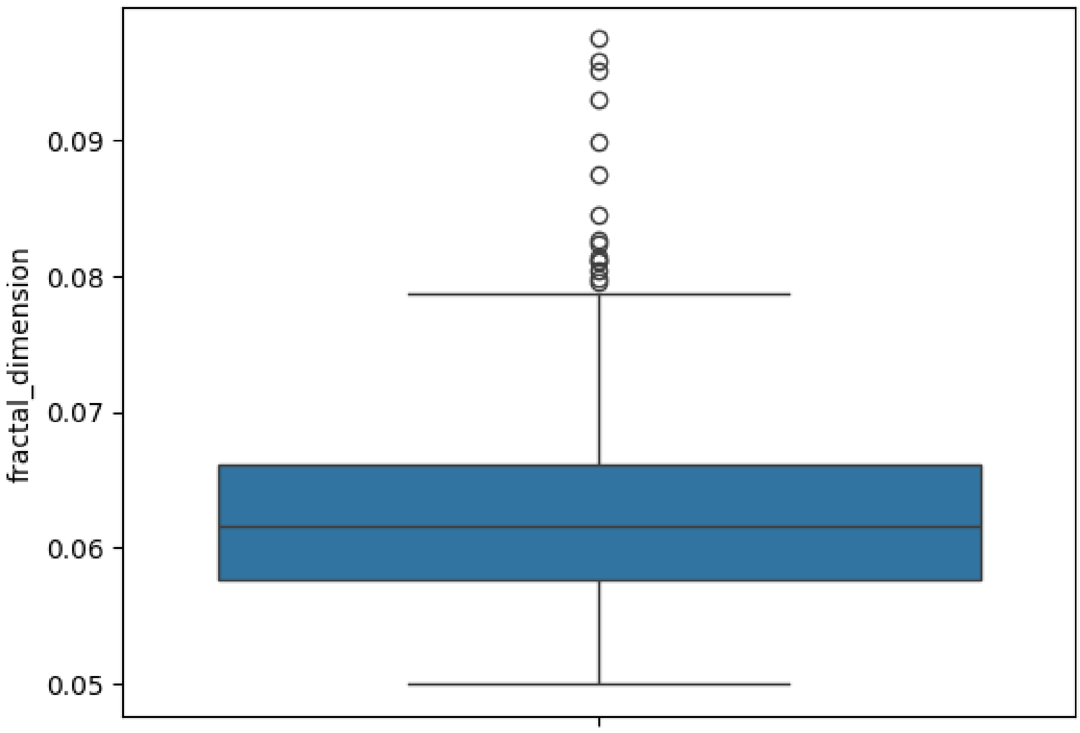

0.16
0.14
0.12
0.10
0.08
0.06
smoothness_mean

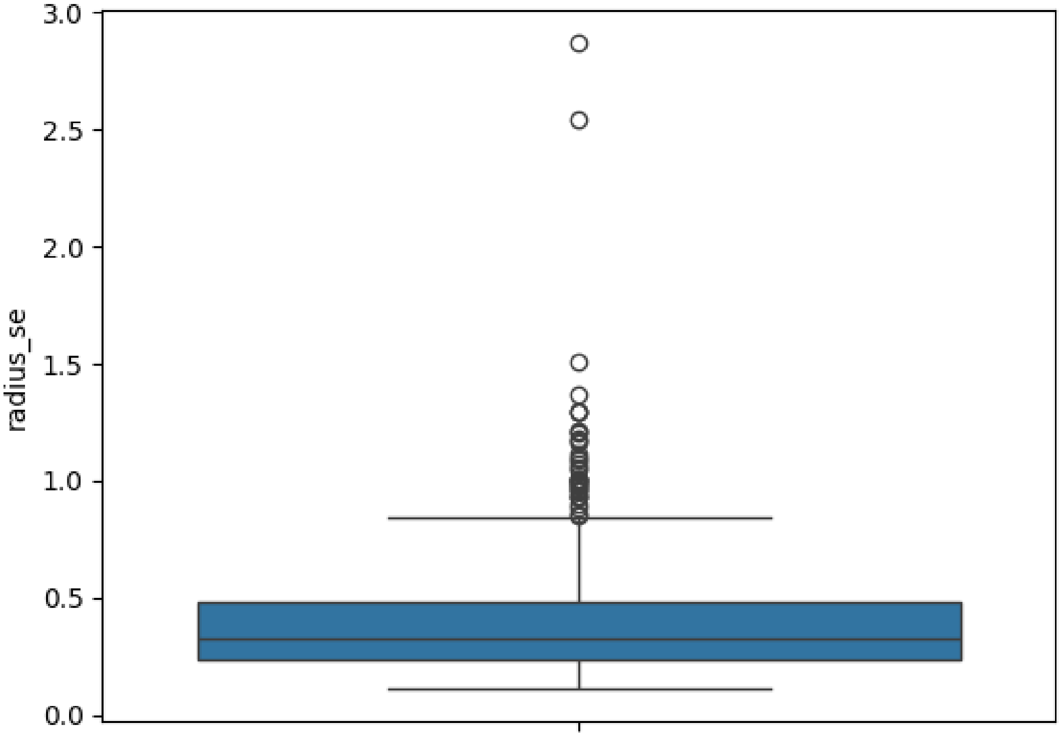

0.35
0.30
0.25
0.20
0.15
0.10
0.05
compactness_mean

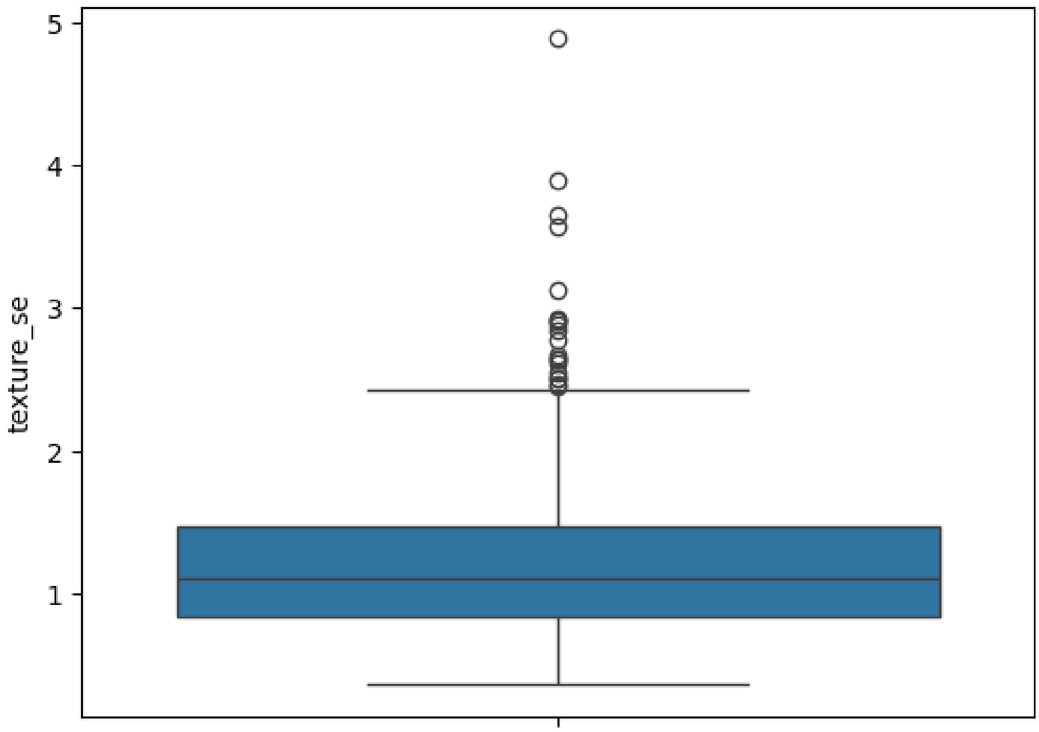

0.4
0.3
0.2
0.1
0.0
concavity_mean

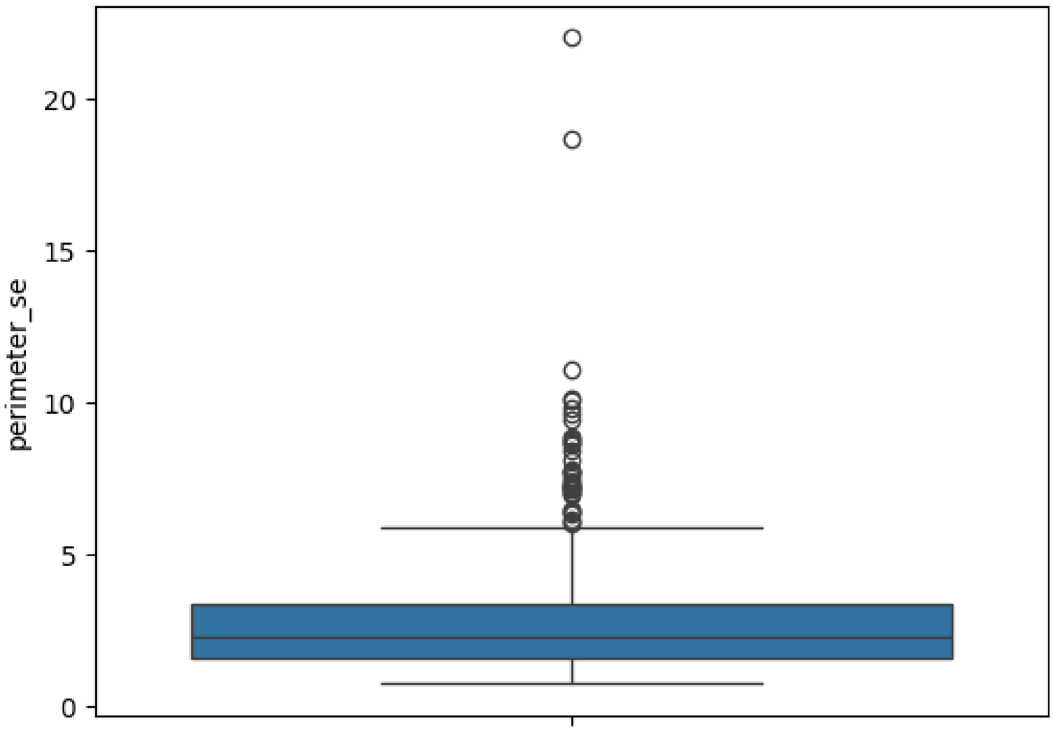

0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
concave_points_mean

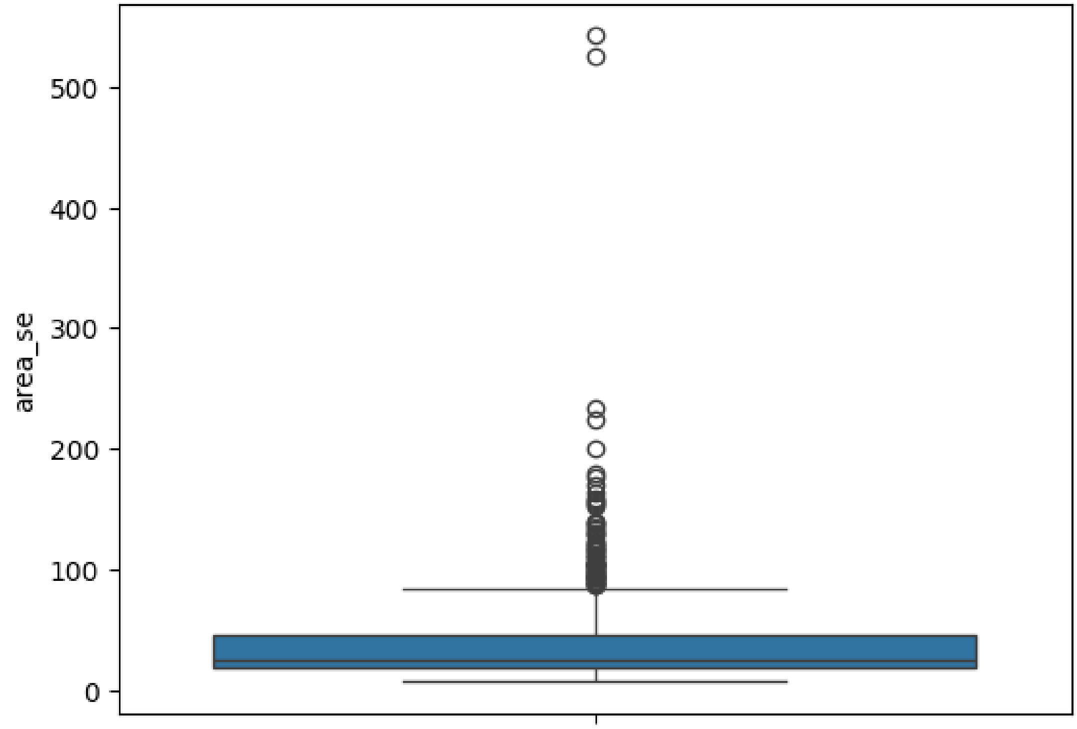

0.300
0.275
0.250
0.225
0.200
0.175
0.150
0.125
0.100
symmetry_mean


0.09
0.08
0.07
0.06
0.05
fractal_dimension

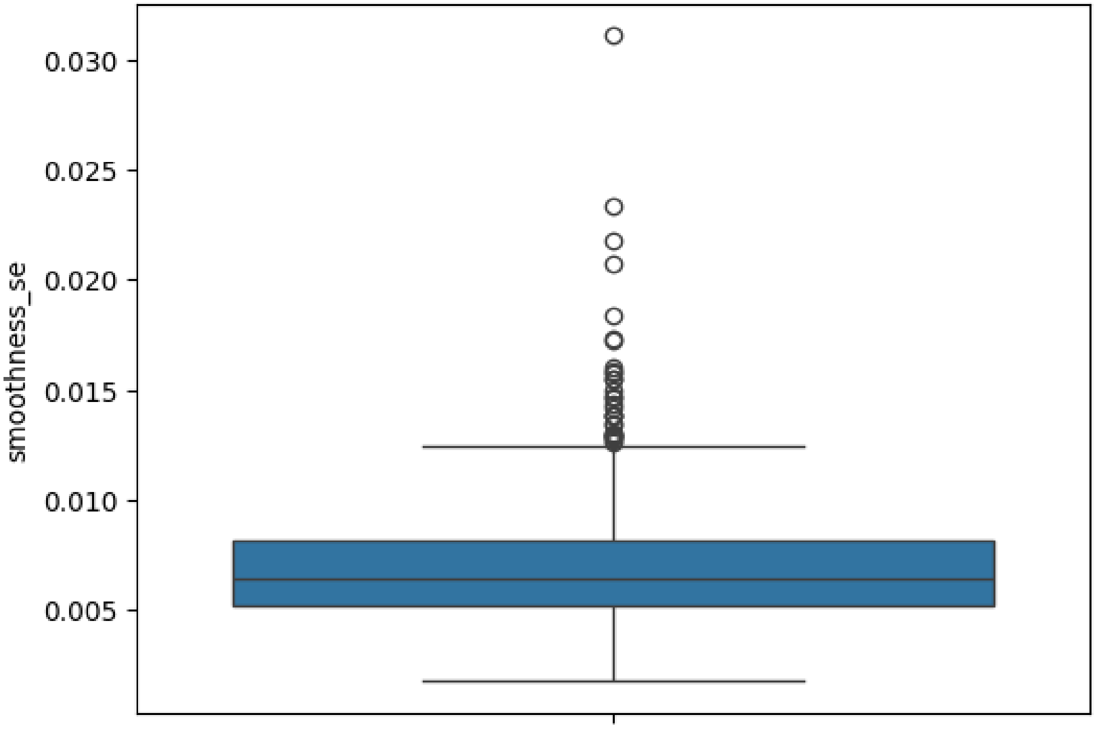

3.0
2.5
2.0
1.5
1.0
0.5
0.0
radius_se


texture_se
5
4
3
2
1

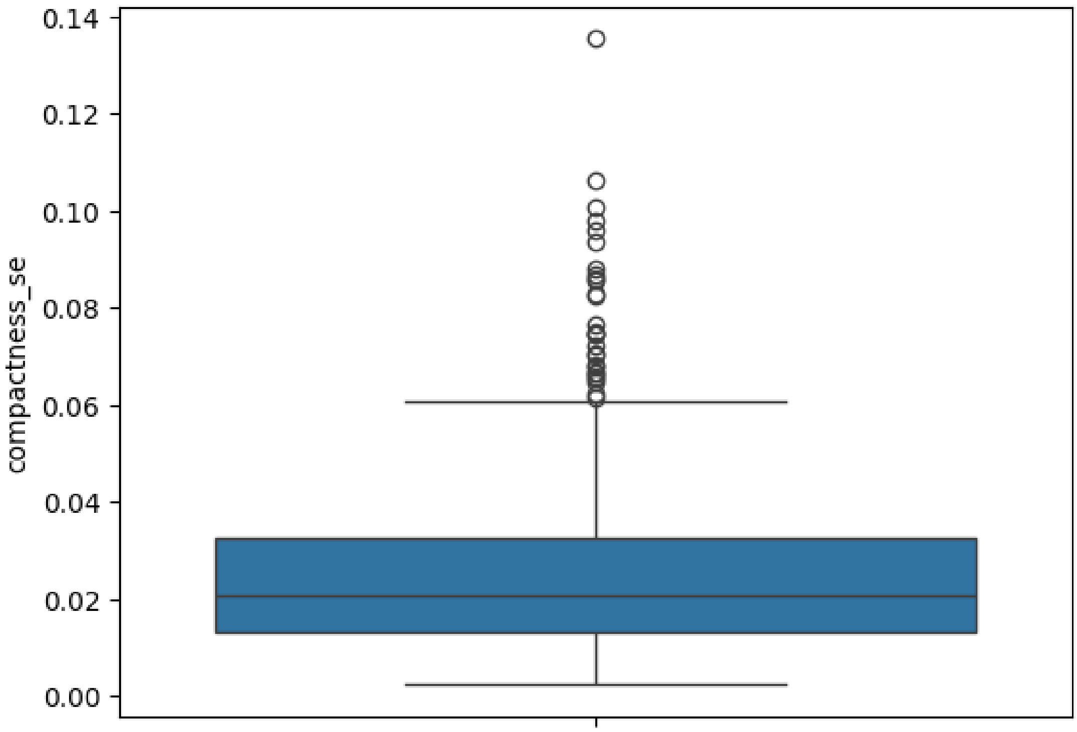

perimeter_se
20
15
10
5
0

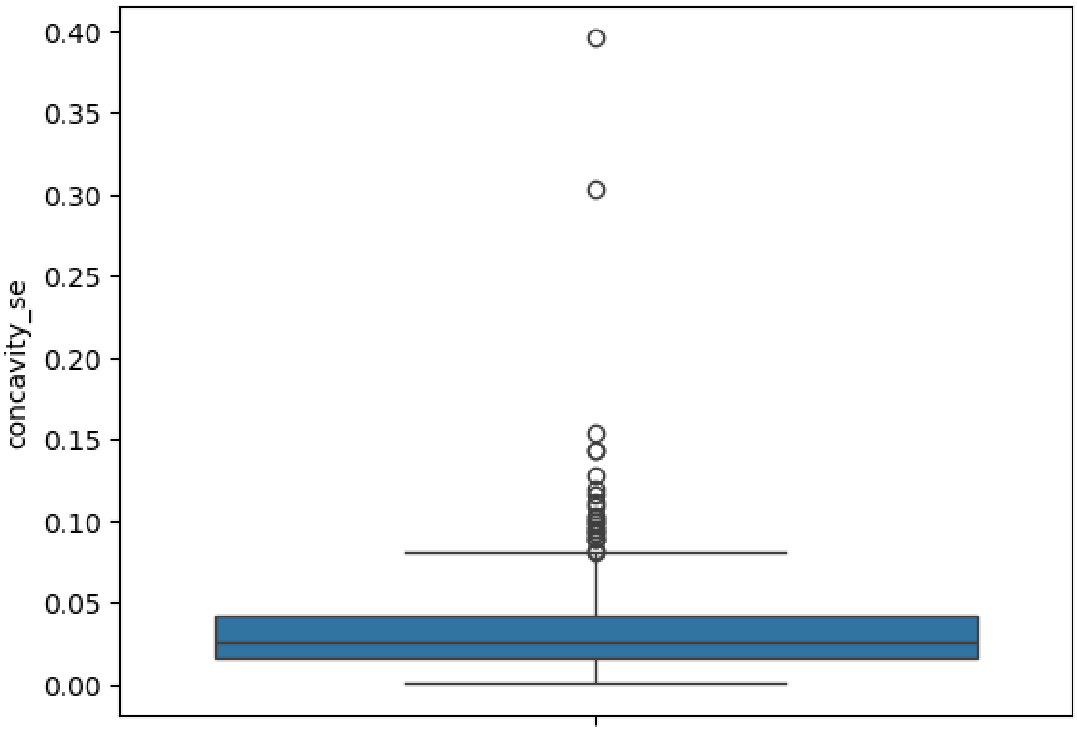


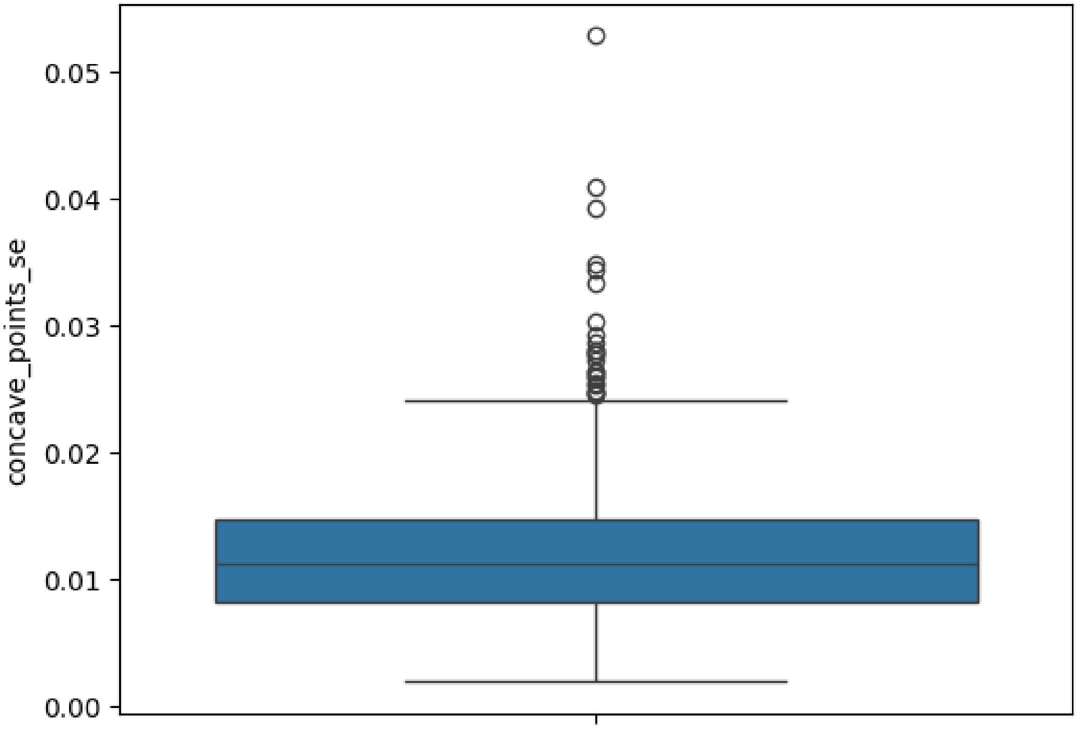
500
400
300
200
100
0
area_se

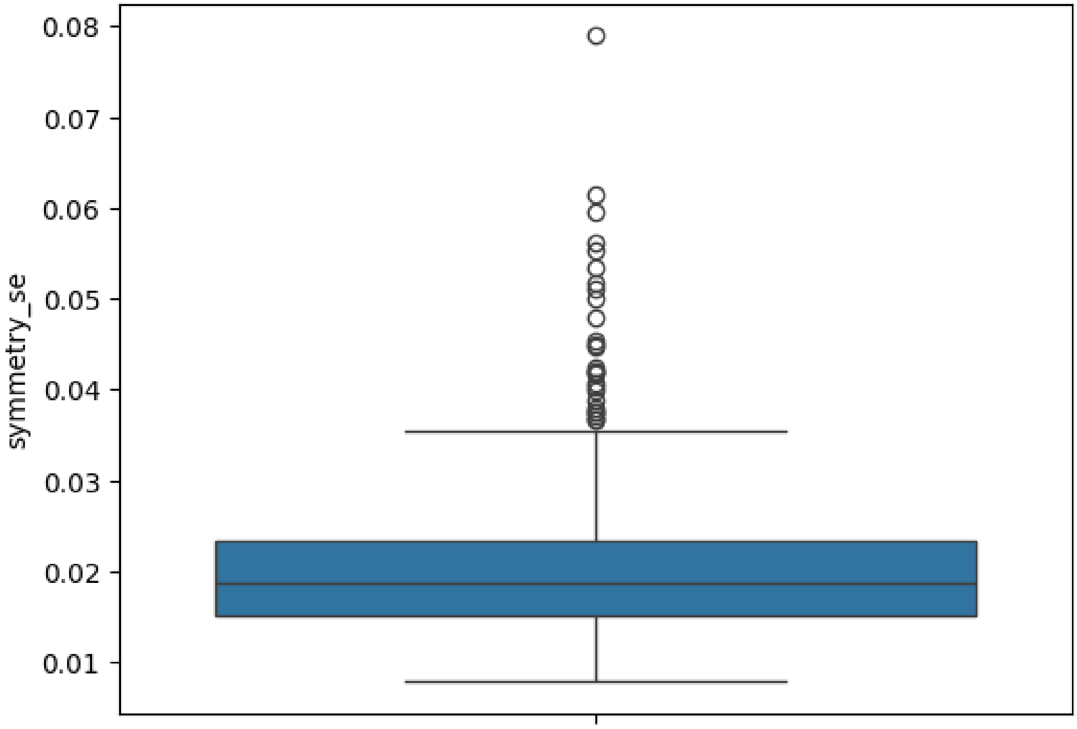


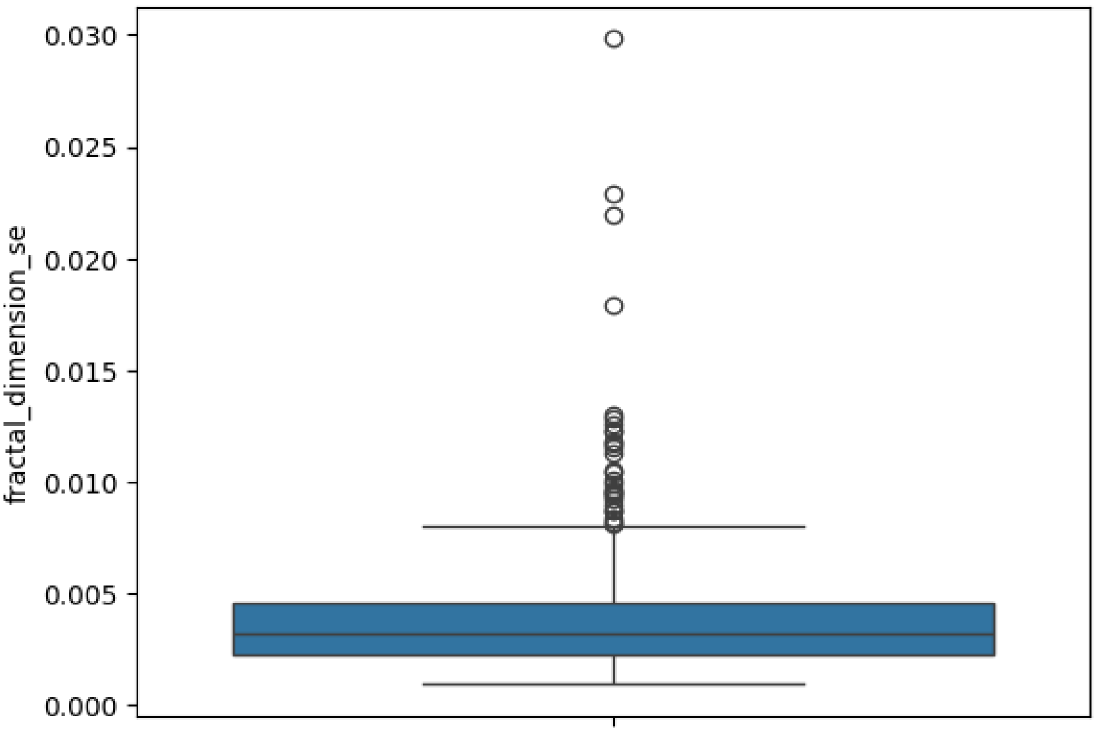
0.030
0.025
0.020
0.015
0.010
0.005
smoothness_se

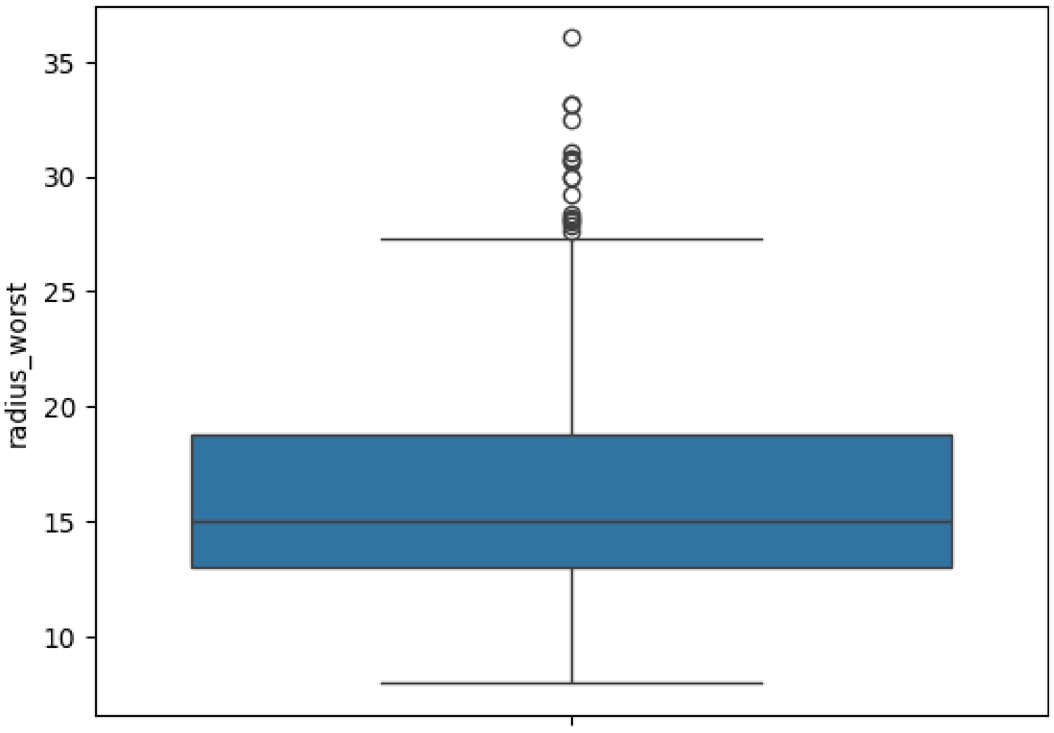

0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
compactness_se

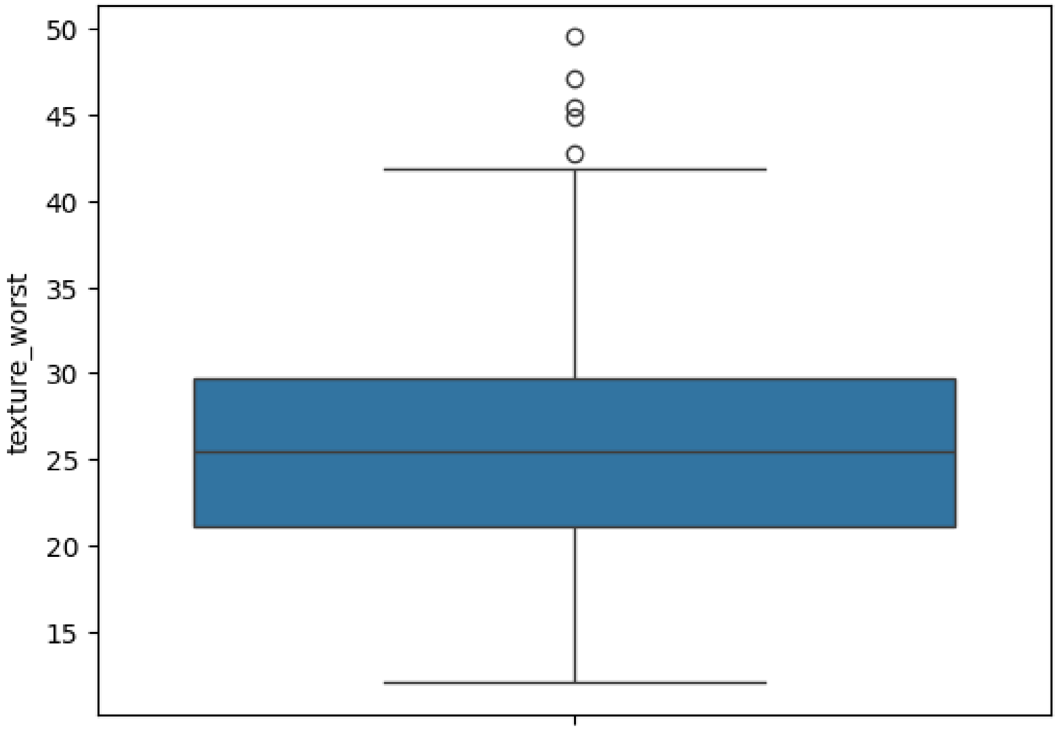

0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
concavity_se

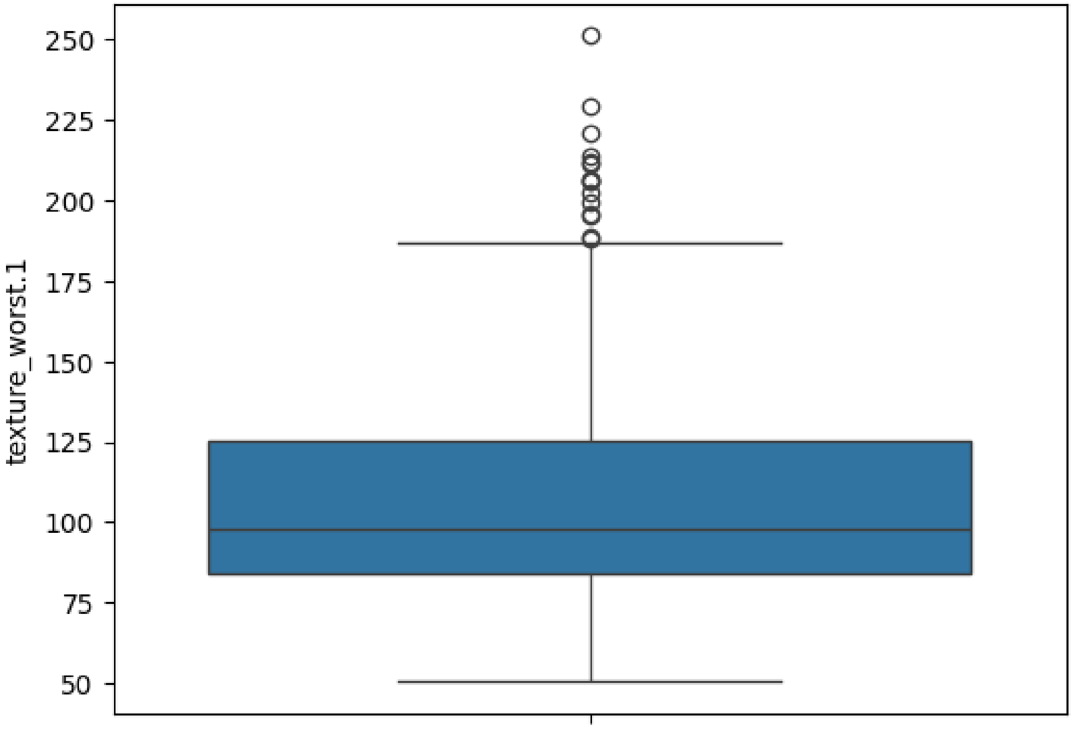

0.05
0.04
0.03
0.02
0.01
0.00
concave_points_se

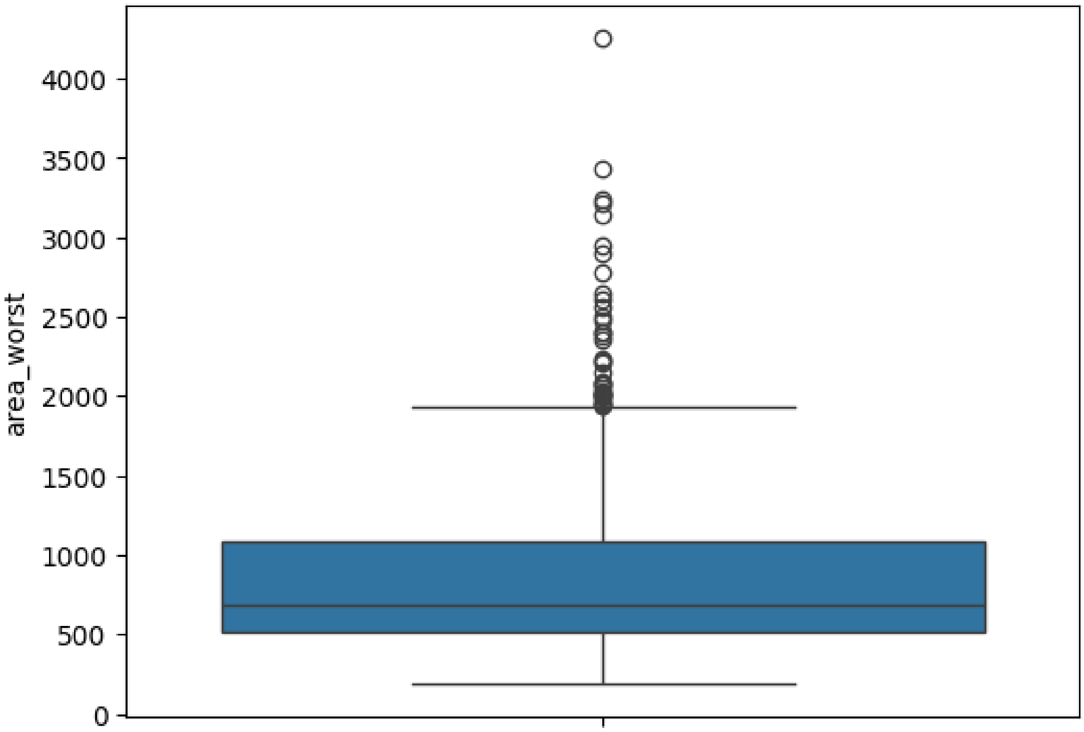

0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
symmetry_se

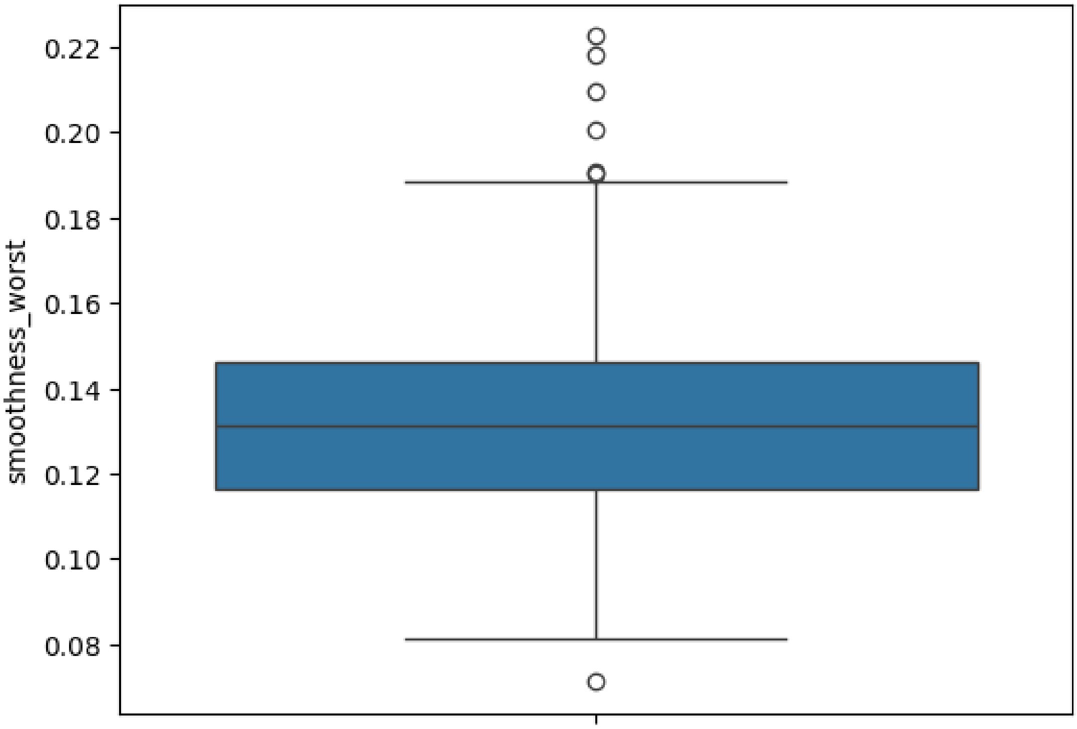

0.030
0.025
0.020
0.015
0.010
0.005
0.000
fractal_dimension_se

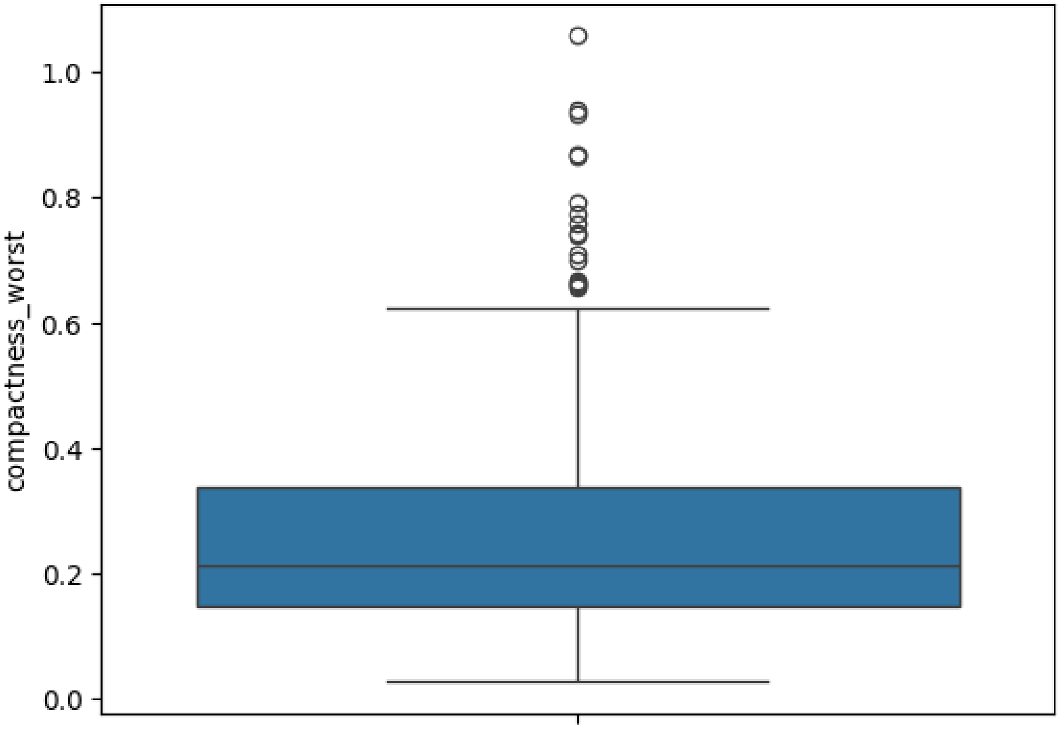

35
30
25
20
15
10
radius_worst

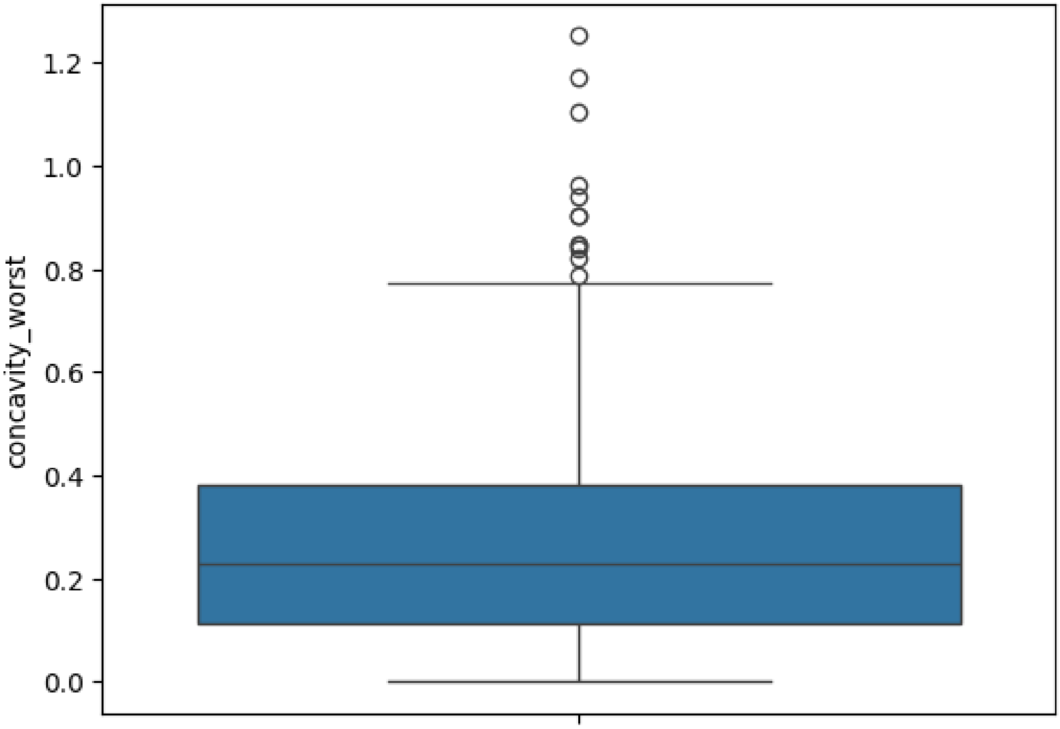

texture_worst
50
45
40
35
30
25
20
15

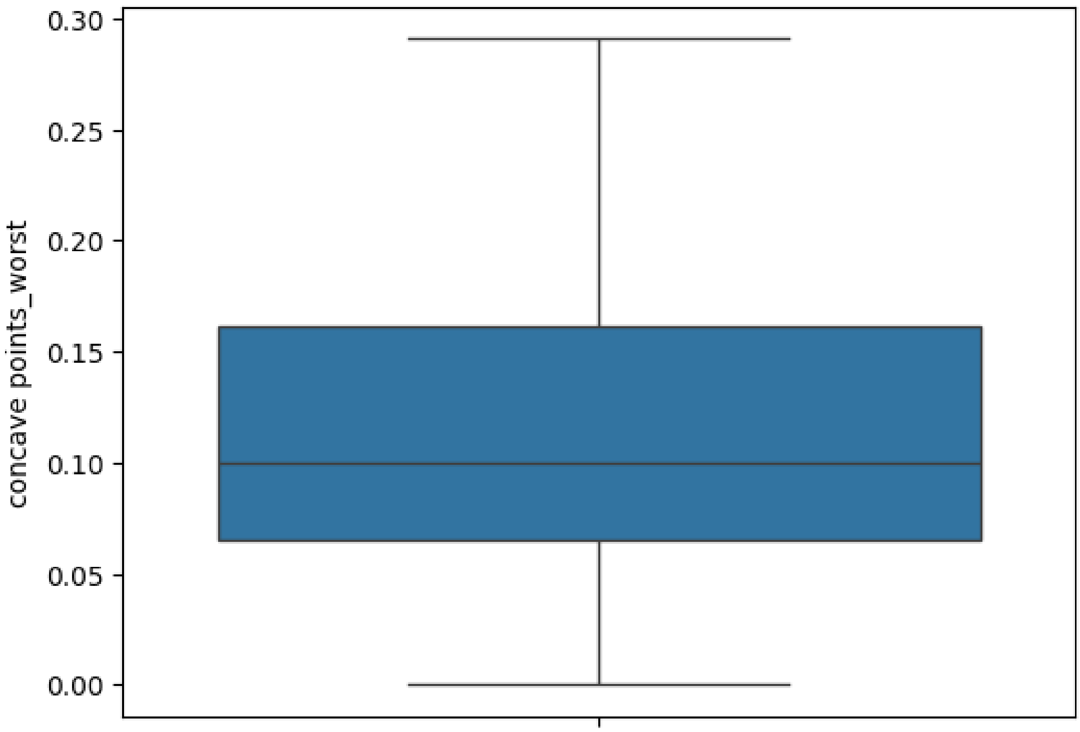

texture_worst.1
250
225
200
175
150
125
100
75
50

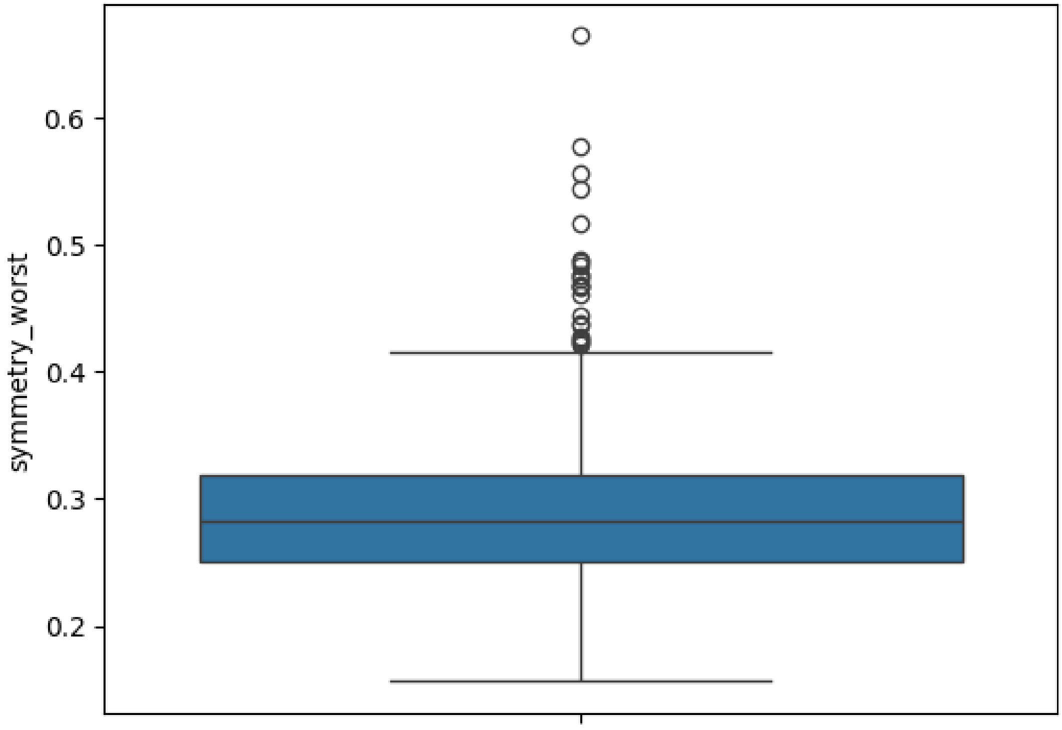

4000
3500
3000
2500
2000
1500
1000
500
0
area_worst

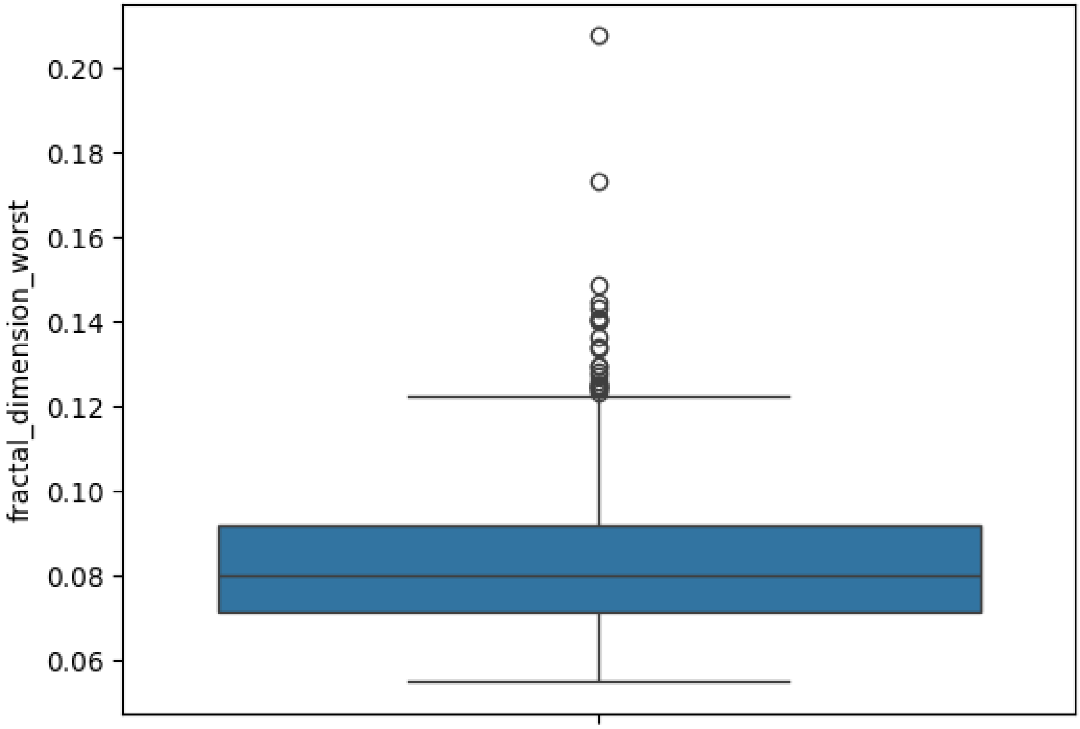

0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
smoothness_worst


compactness_worst
1.0
0.8
0.6
0.4
0.2
0.0


concavity_worst
1.2
1.0
0.8
0.6
0.4
0.2
0.0

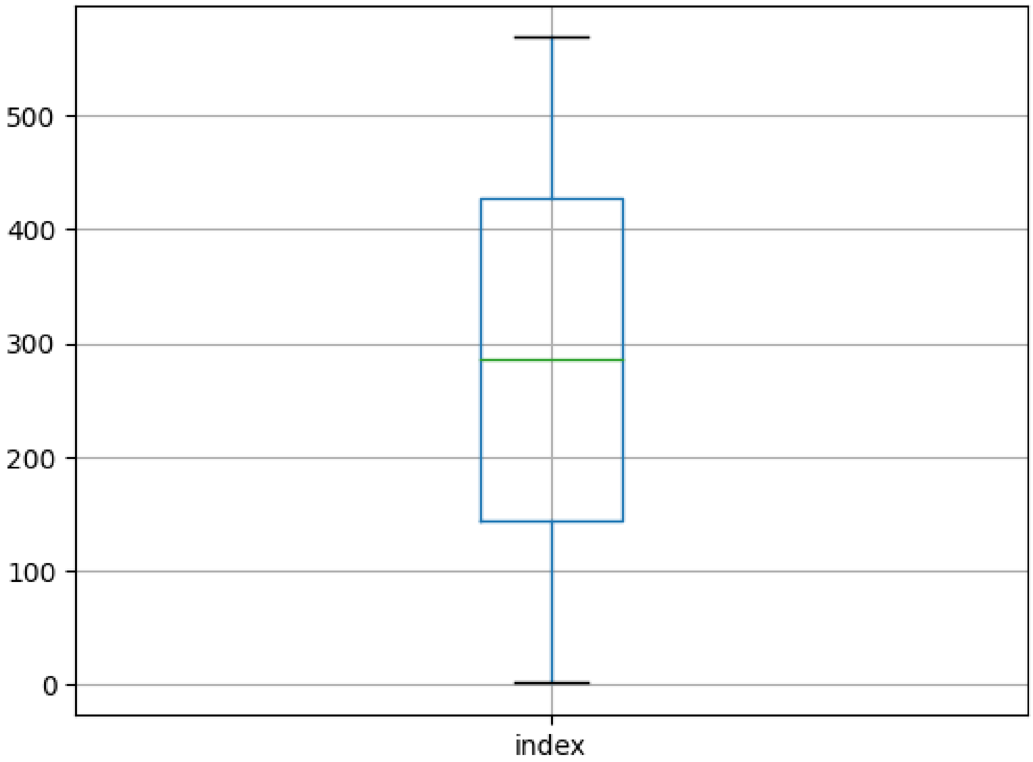

concave points_worst
0.30
0.25
0.20
0.15
0.10
0.05
0.00

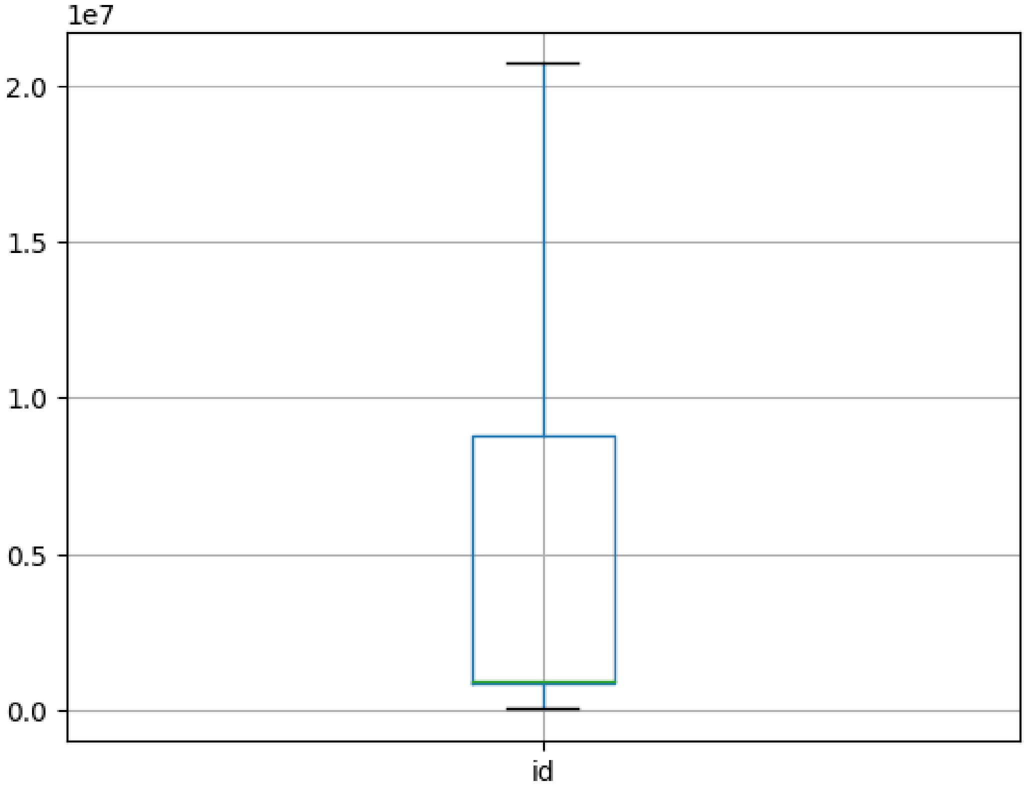

symmetry_worst
0.6
0.5
0.4
0.3
0.2

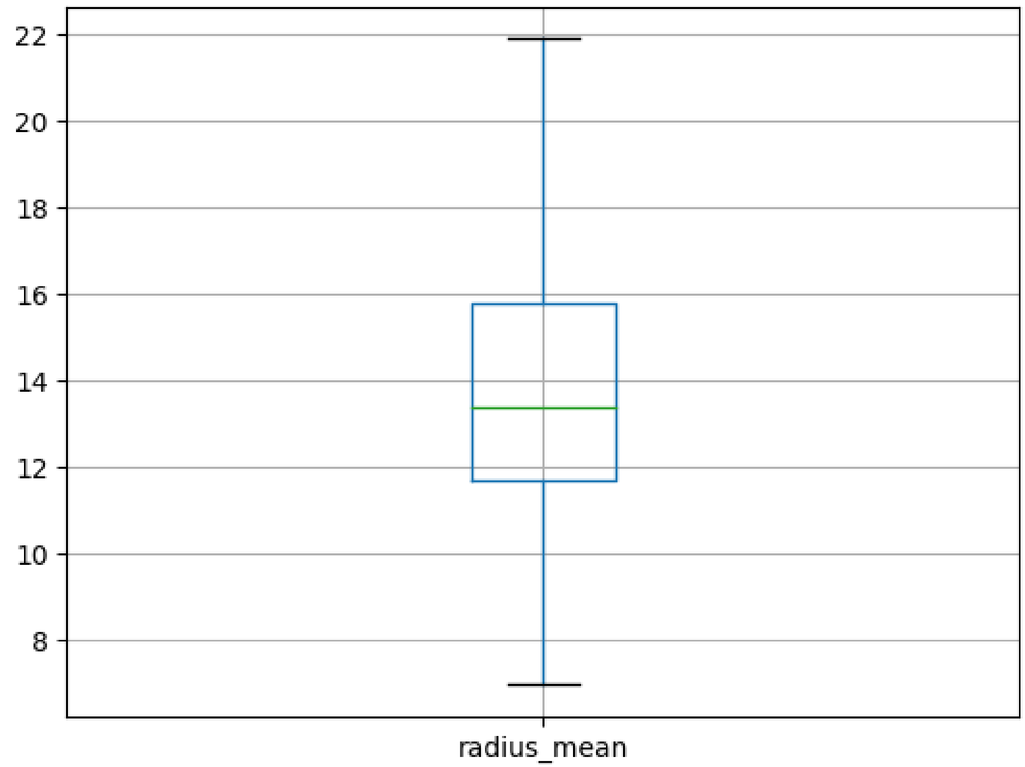


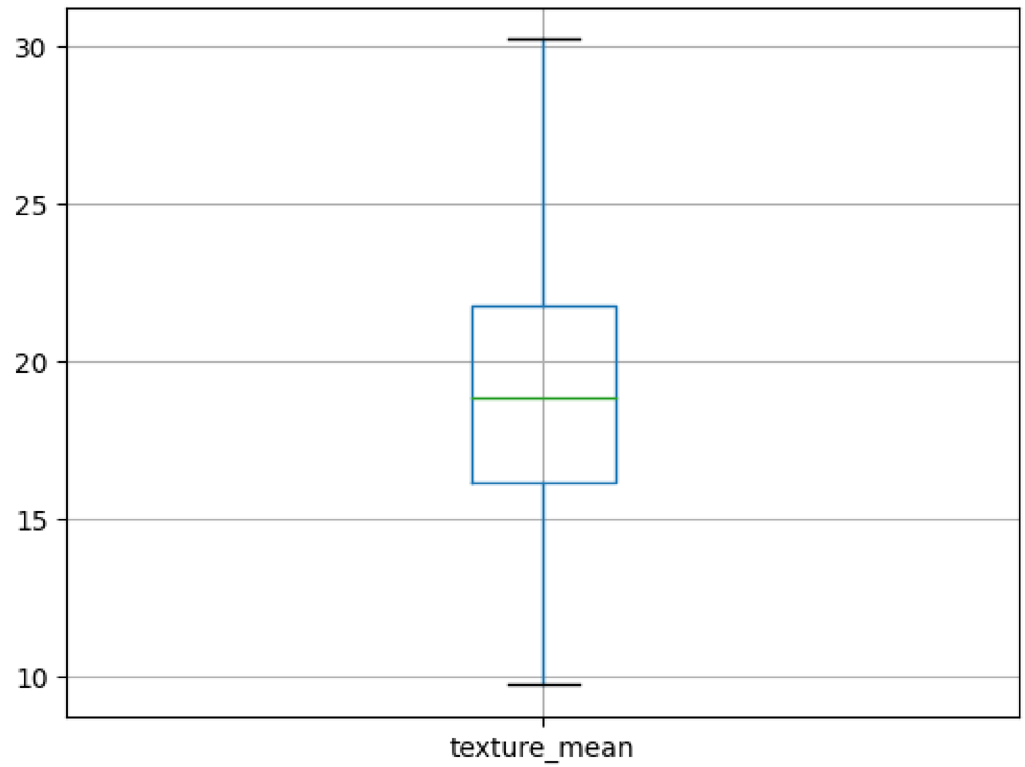
[20]:
```
for column in numerical_cols:
  Q1 = cancer_data[column].quantile(.25)
  Q3 = cancer_data[column].quantile(.75)
  IQR = Q3 - Q1
  UB =  Q3 + 1.5 * IQR
  LB =  Q1 - 1.5 * IQR
  cancer_data.loc[cancer_data[column] >=UB, column]=UB
  cancer_data.loc[cancer_data[column] <=LB, column]=LB
  plt.figure()
  cancer_data[[column]].boxplot()
```

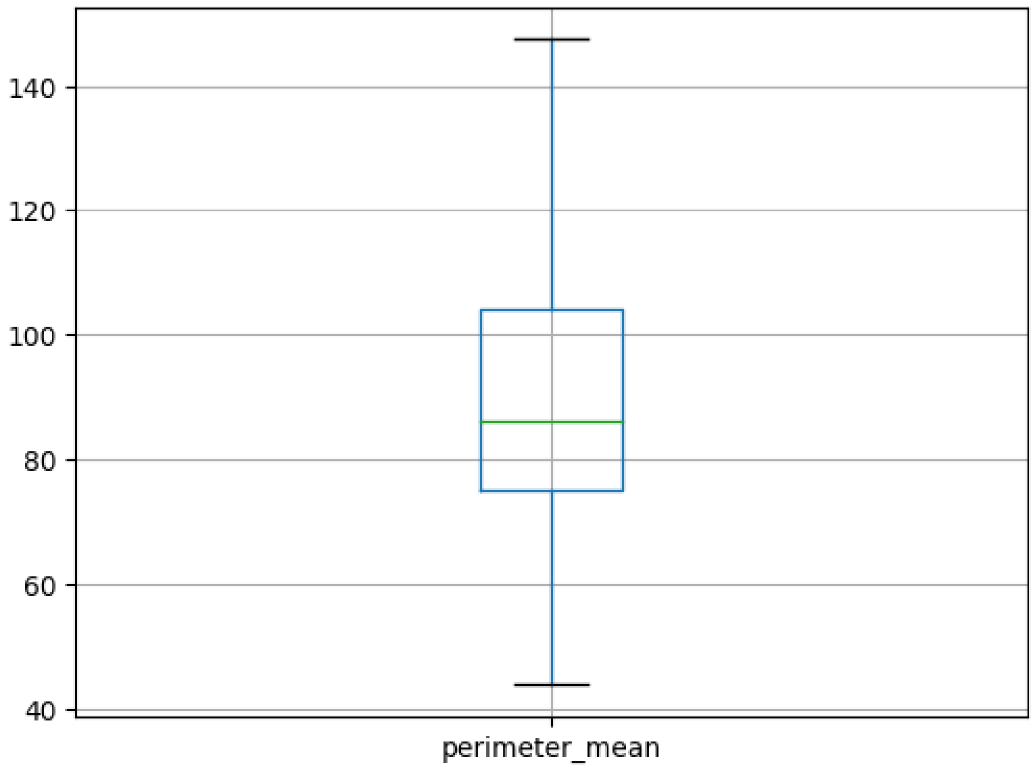
```
<ipython-input-20-0e36a88e36ee>:9: RuntimeWarning: More than 20 figures have
been opened. Figures created through the pyplot interface
(`matplotlib.pyplot.figure`) are retained until explicitly closed and may
consume too much memory. (To control this warning, see the rcParam
`figure.max_open_warning`). Consider using `matplotlib.pyplot.close()`.
  plt.figure()
```

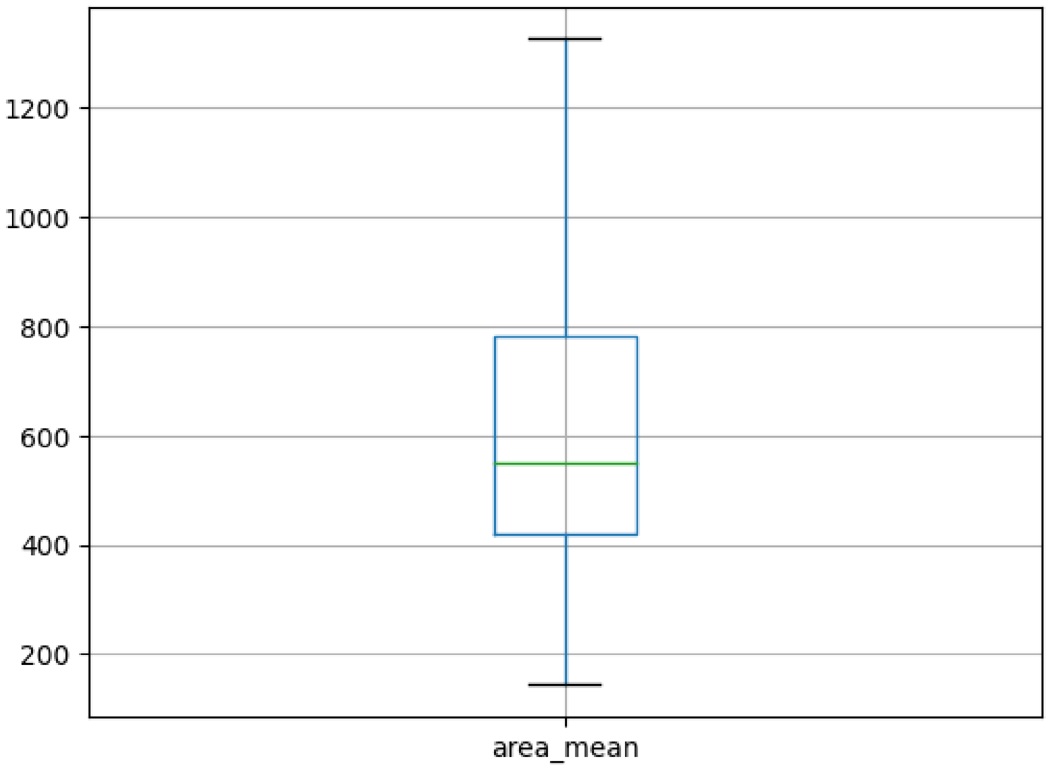

500
400
300
200
100
0
index

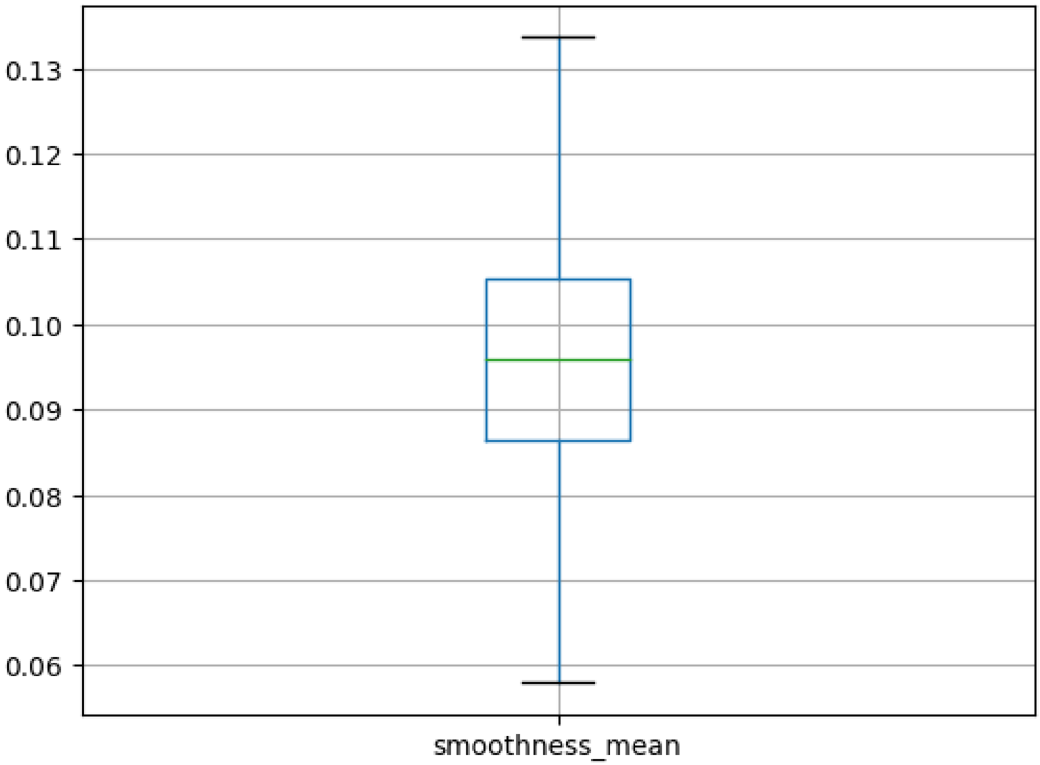

1e7
2.0
1.5
1.0
0.5
0.0
id

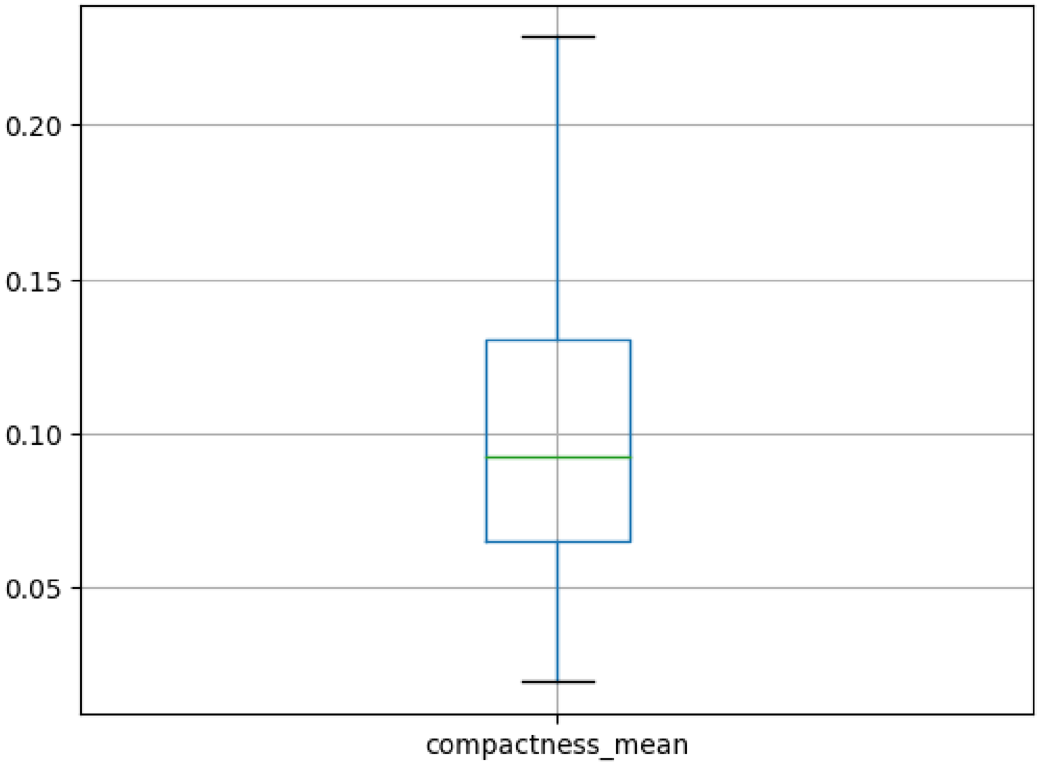

22
20
18
16
14
12
10
8
radius_mean

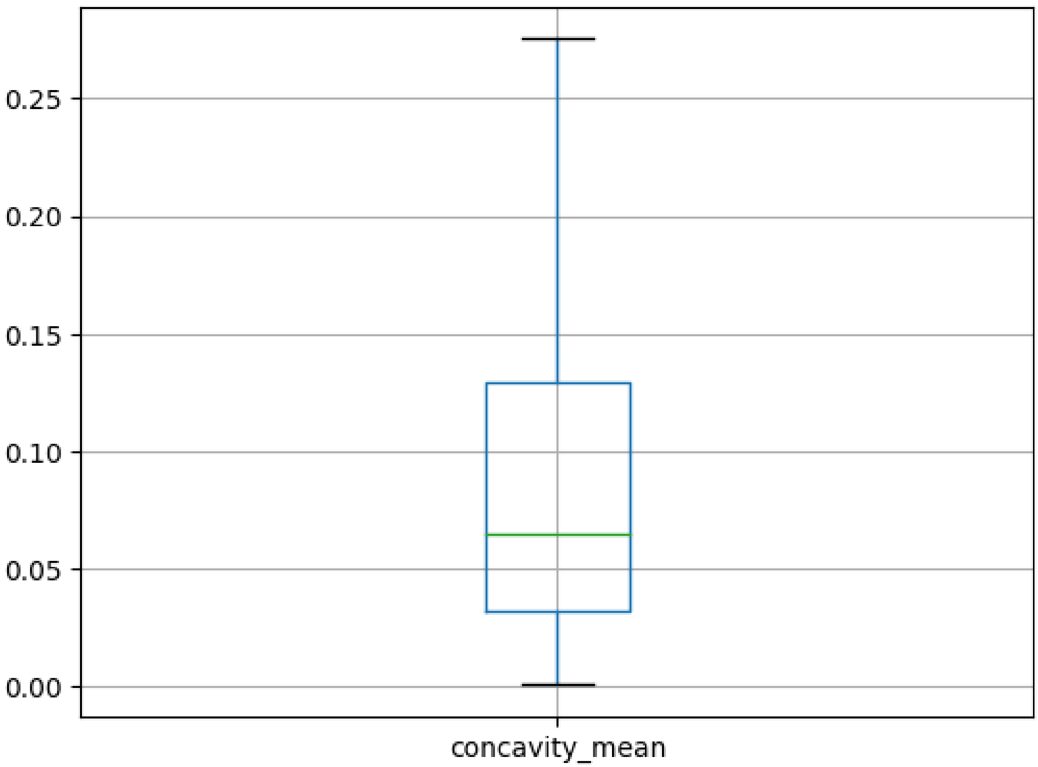

30
25
20
15
10
texture_mean

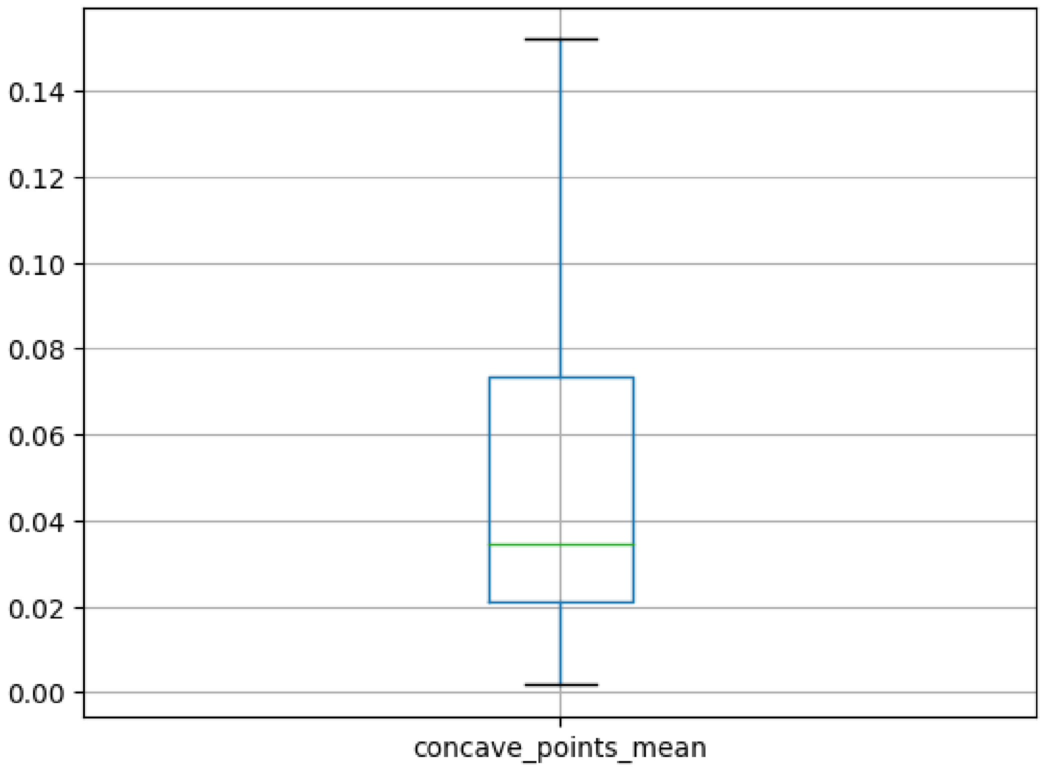

140
120
100
80
60
40
perimeter_mean

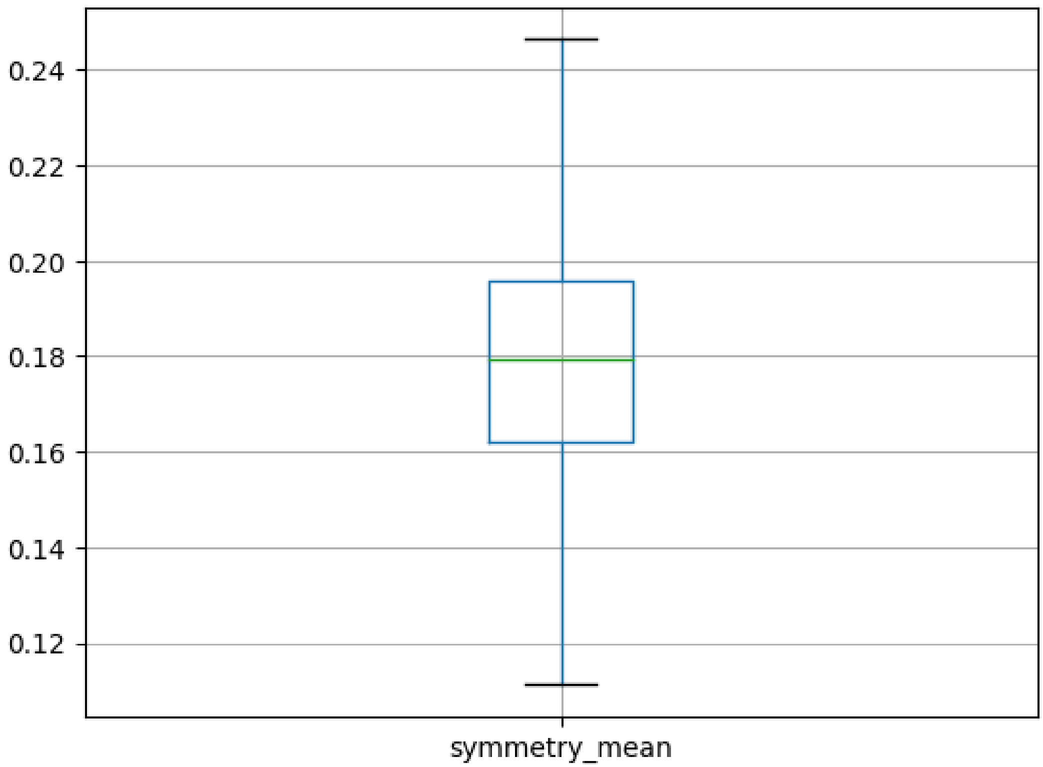

1200
1000
800
600
400
200
area_mean

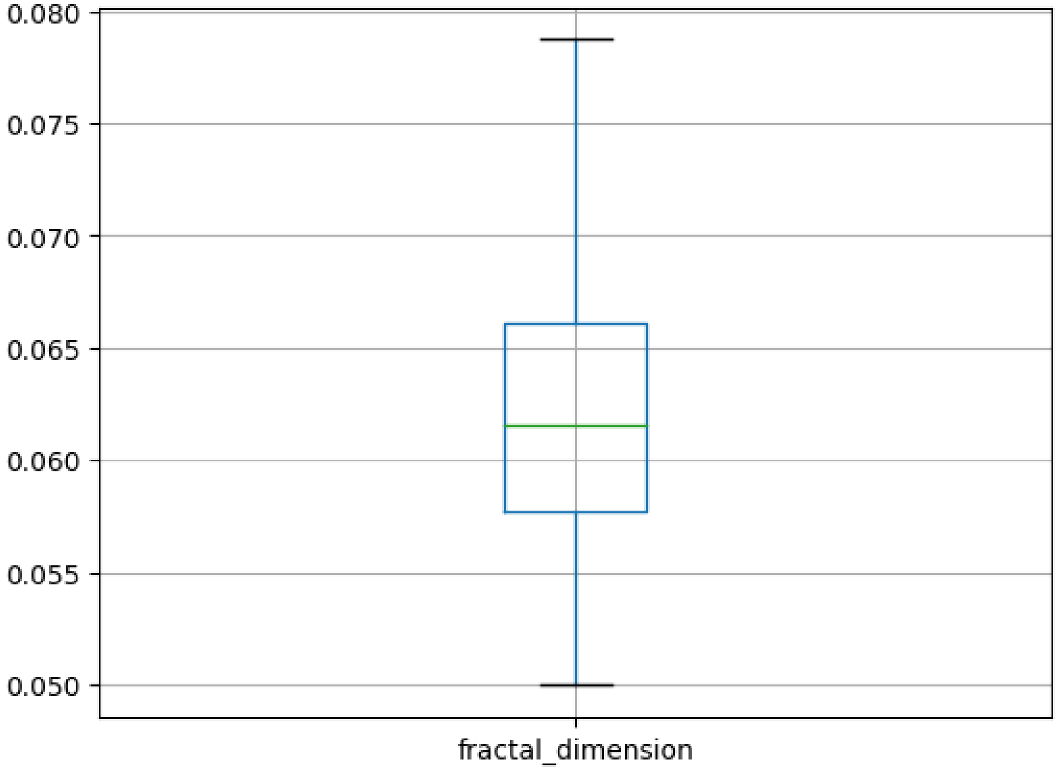

0.13
0.12
0.11
0.10
0.09
0.08
0.07
0.06
smoothness_mean

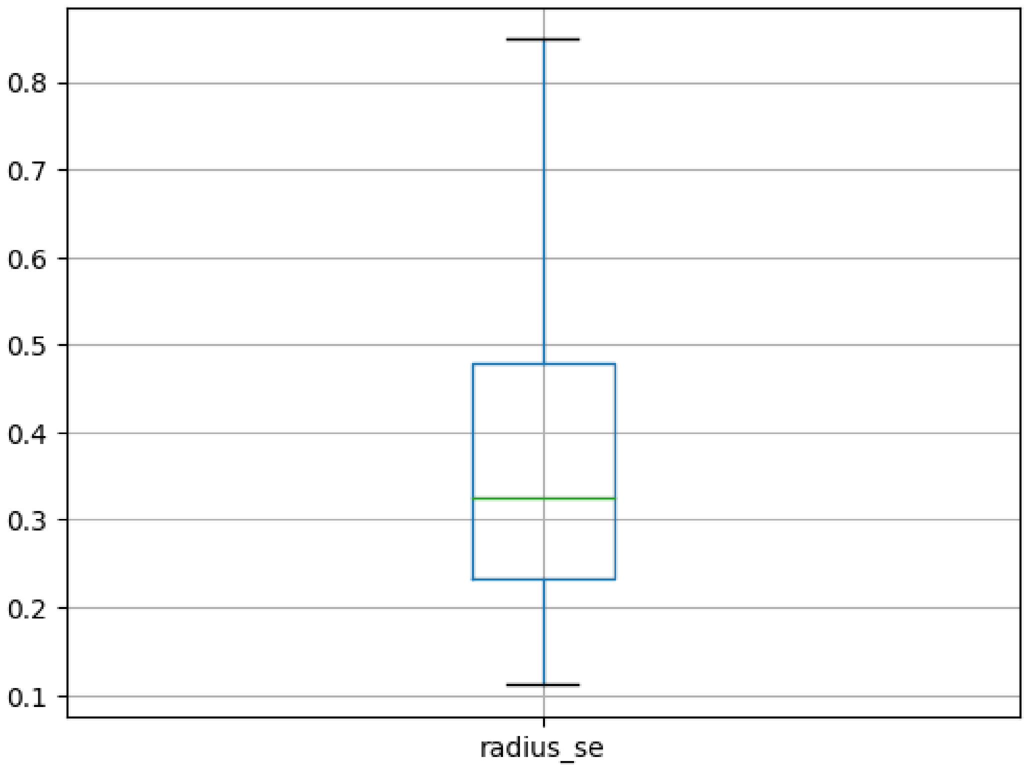

0.20
0.15
0.10
0.05
compactness_mean

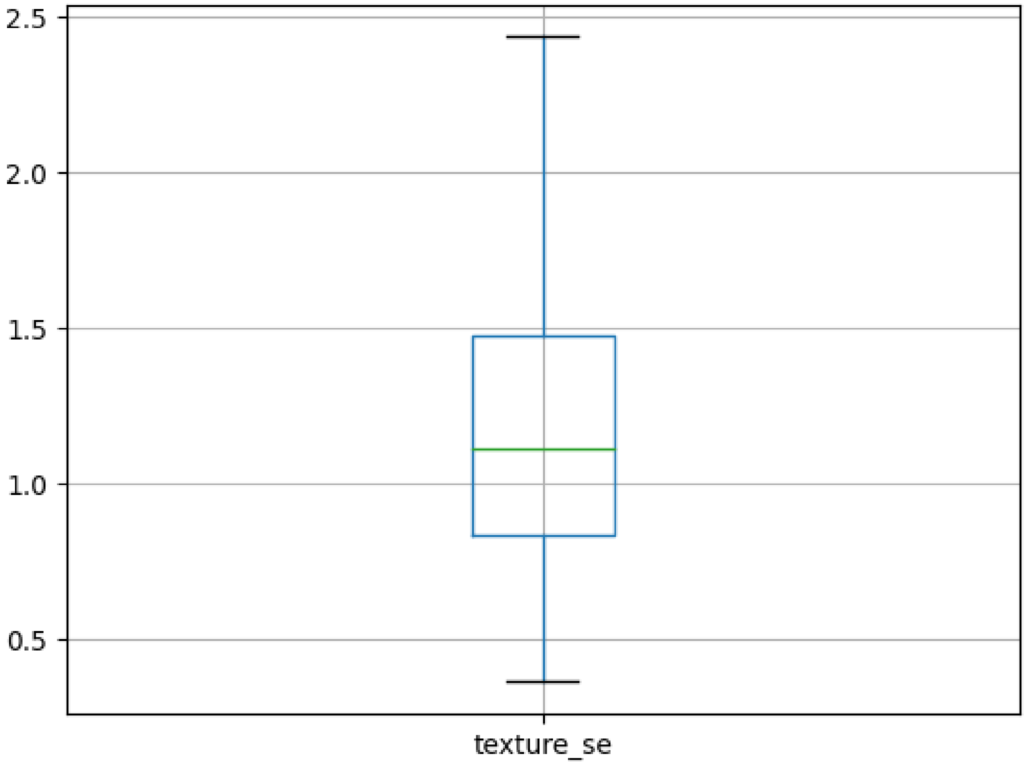

0.25
0.20
0.15
0.10
0.05
0.00
concavity_mean

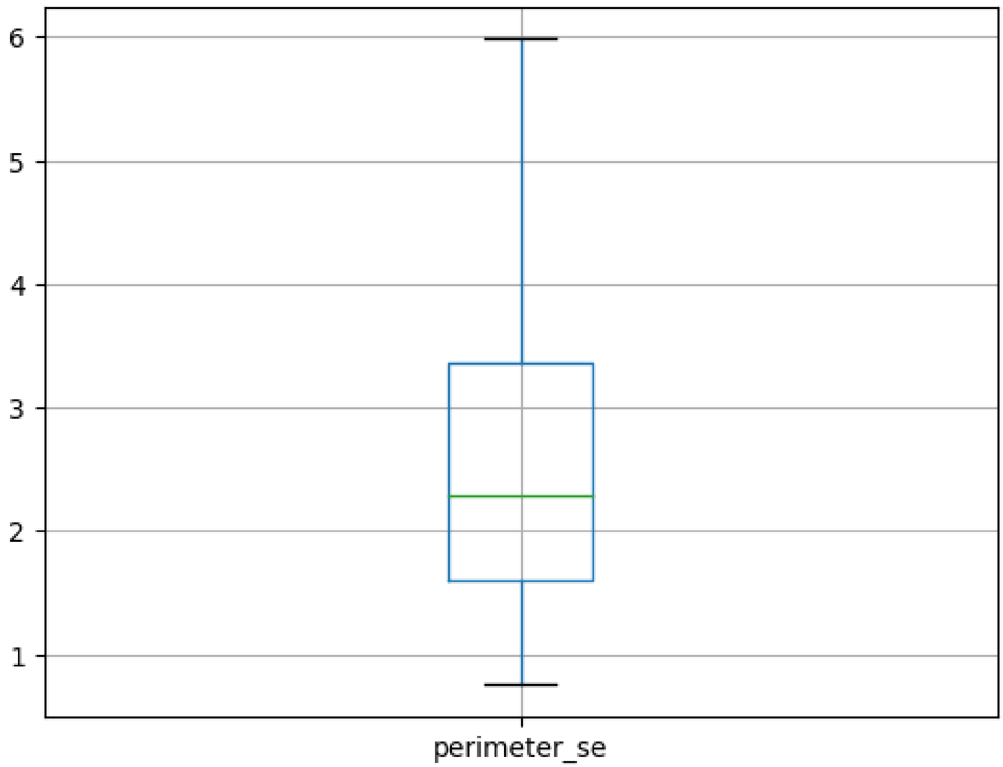

0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
concave_points_mean

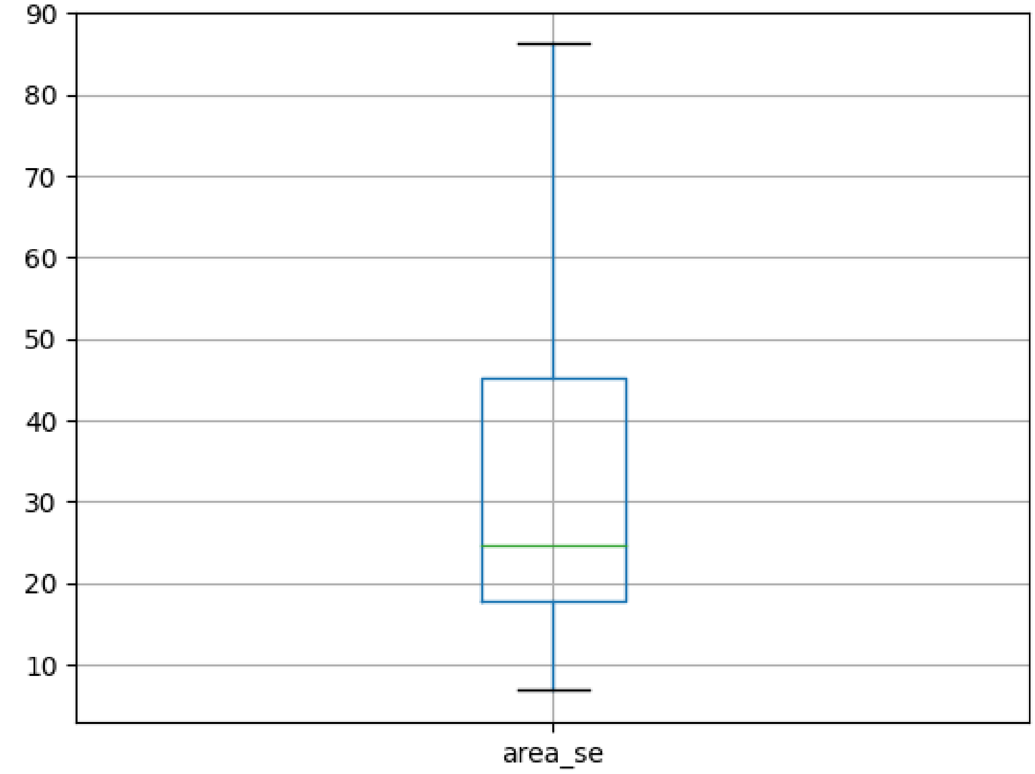

0.24
0.22
0.20
0.18
0.16
0.14
0.12
symmetry_mean

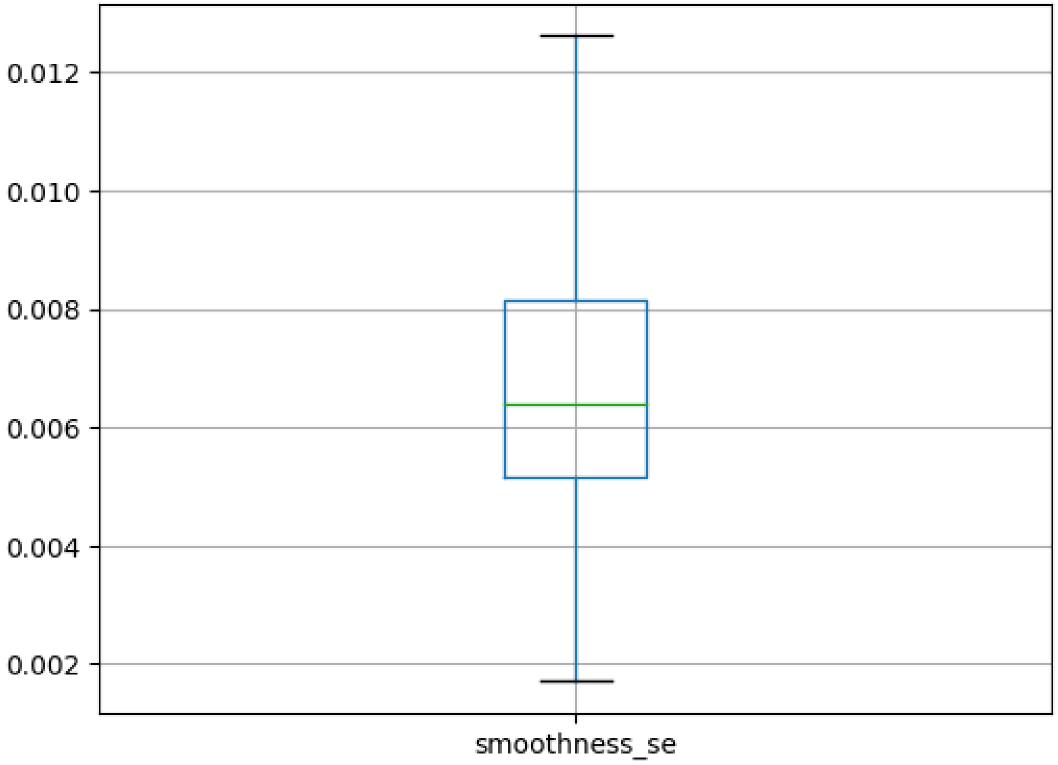

0.080
0.075
0.070
0.065
0.060
0.055
0.050
fractal_dimension

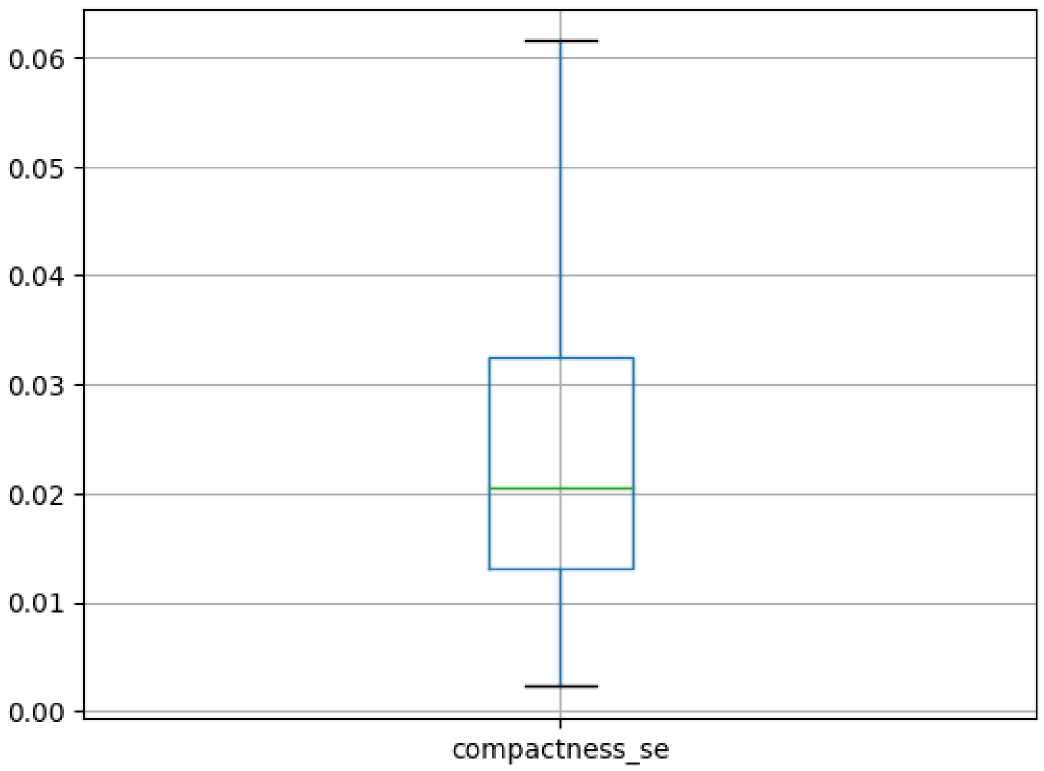

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
radius_se

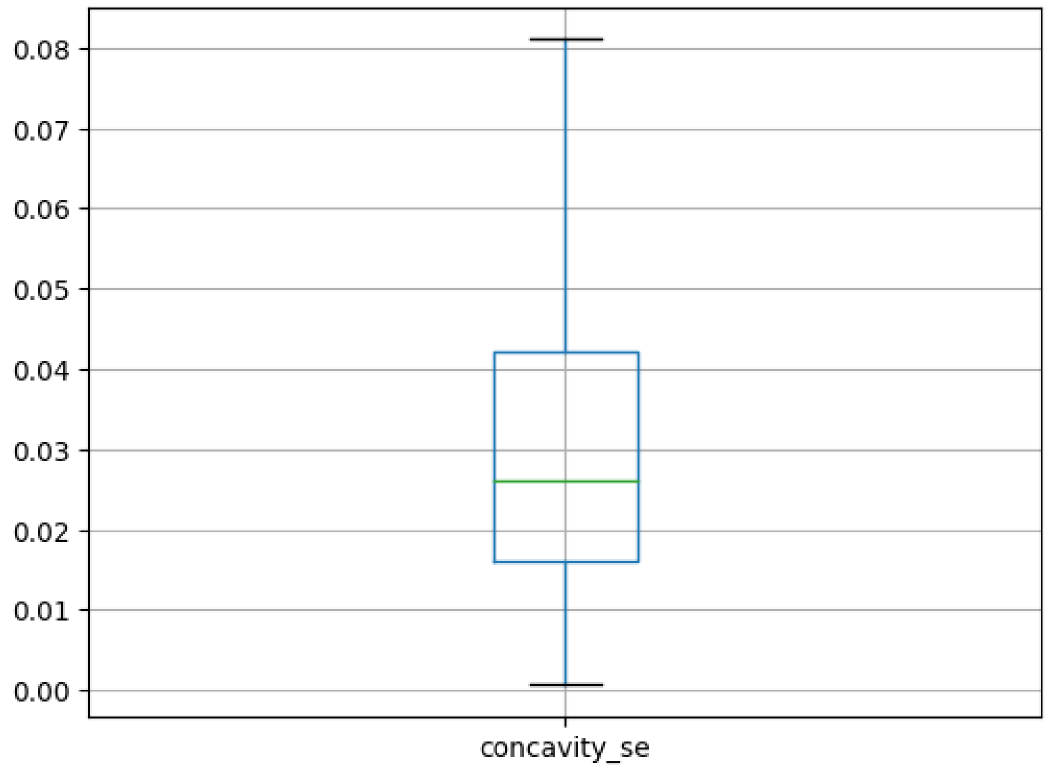

2.5
2.0
1.5
1.0
0.5
texture_se

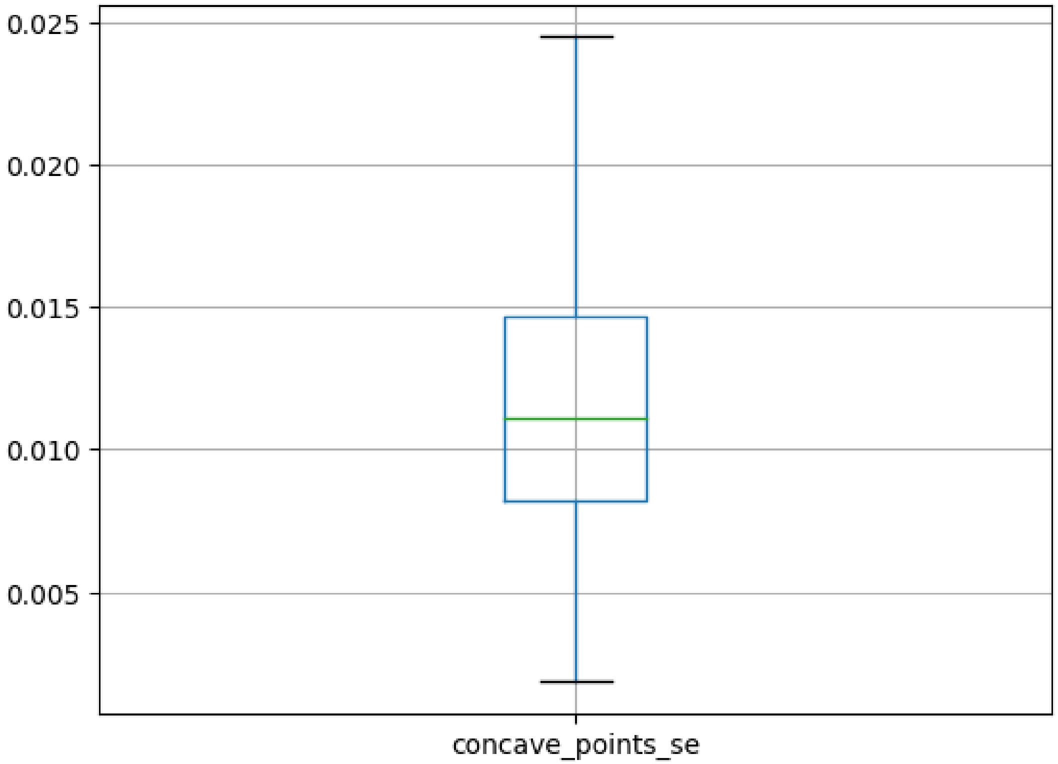

6
5
4
3
2
1
perimeter_se

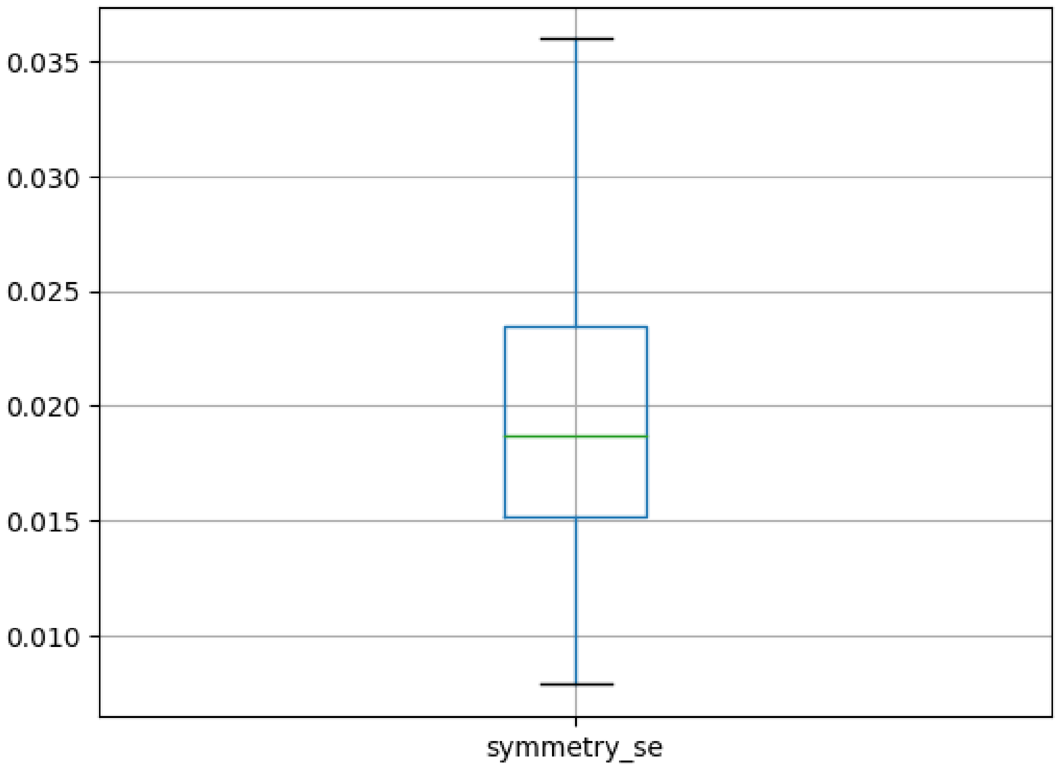

90
80
70
60
50
40
30
20
10
area_se

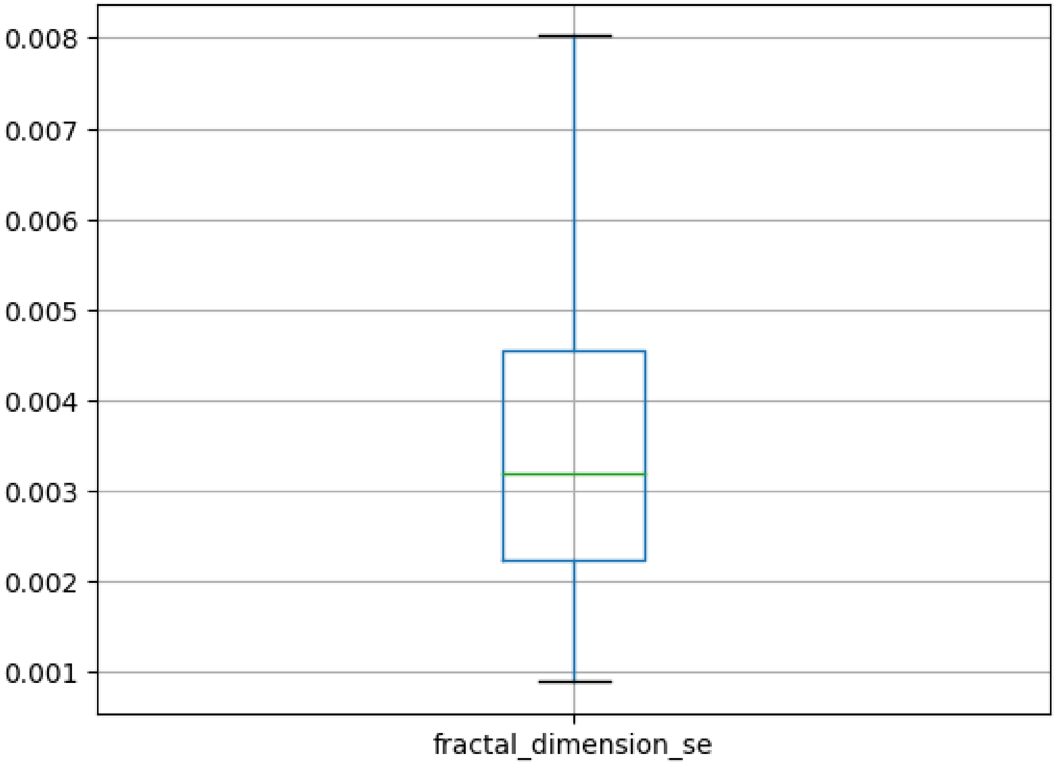

0.012
0.010
0.008
0.006
0.004
0.002
smoothness_se

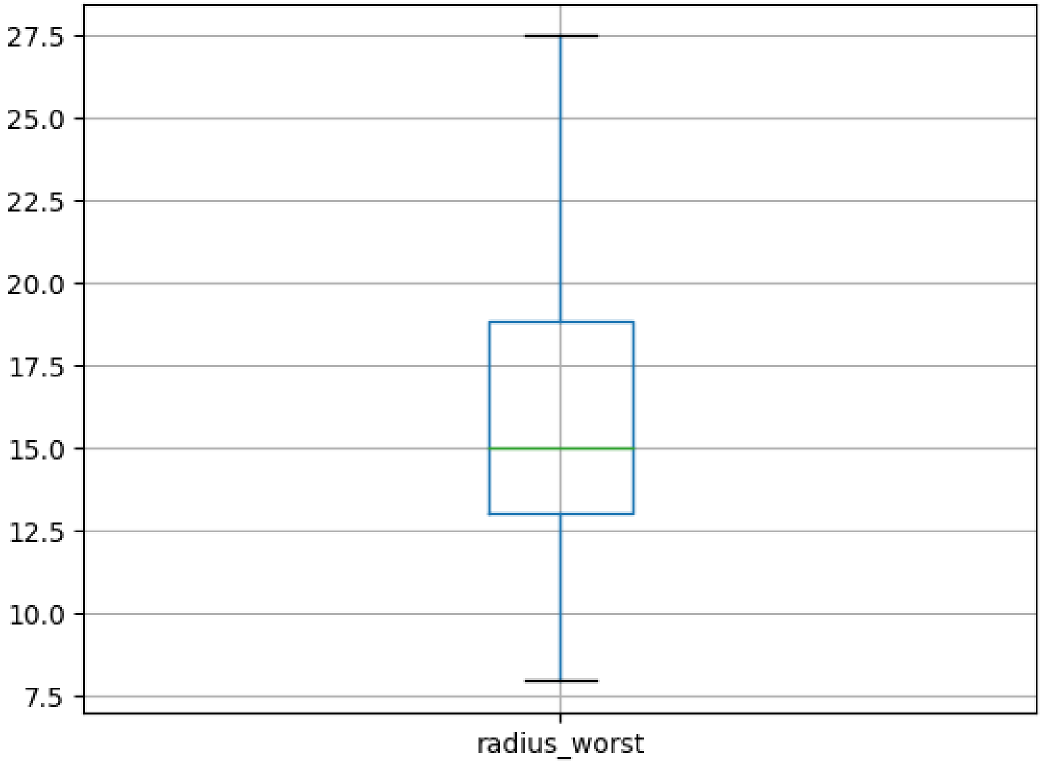

0.06
0.05
0.04
0.03
0.02
0.01
0.00
compactness_se

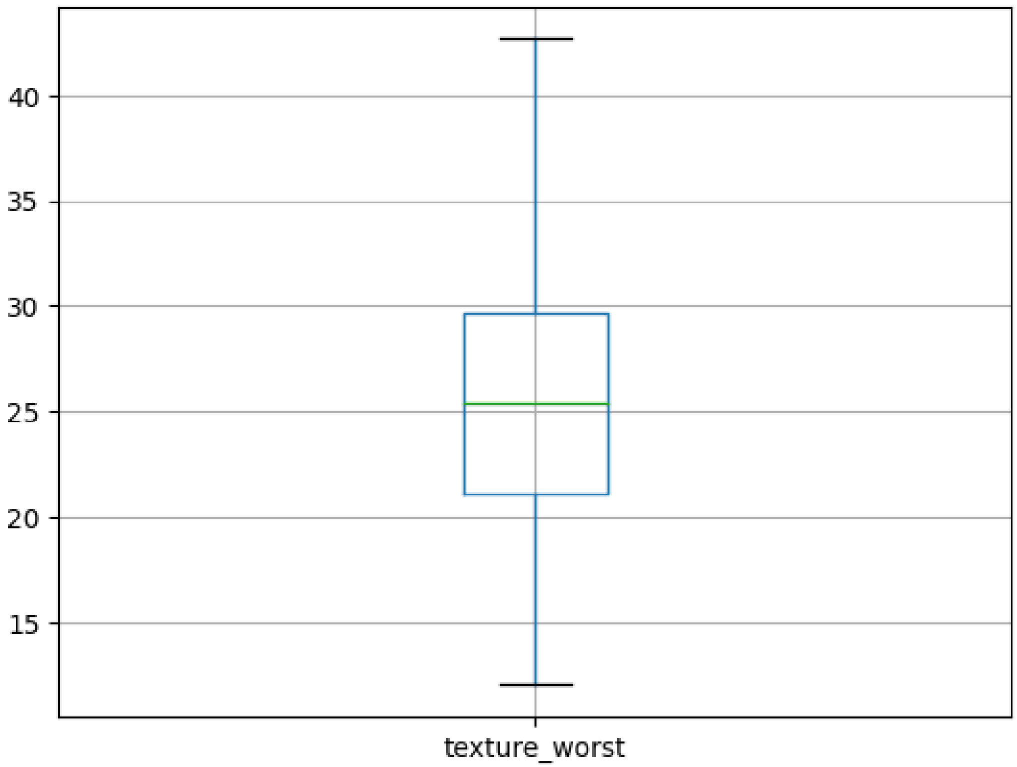

0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
concavity_se

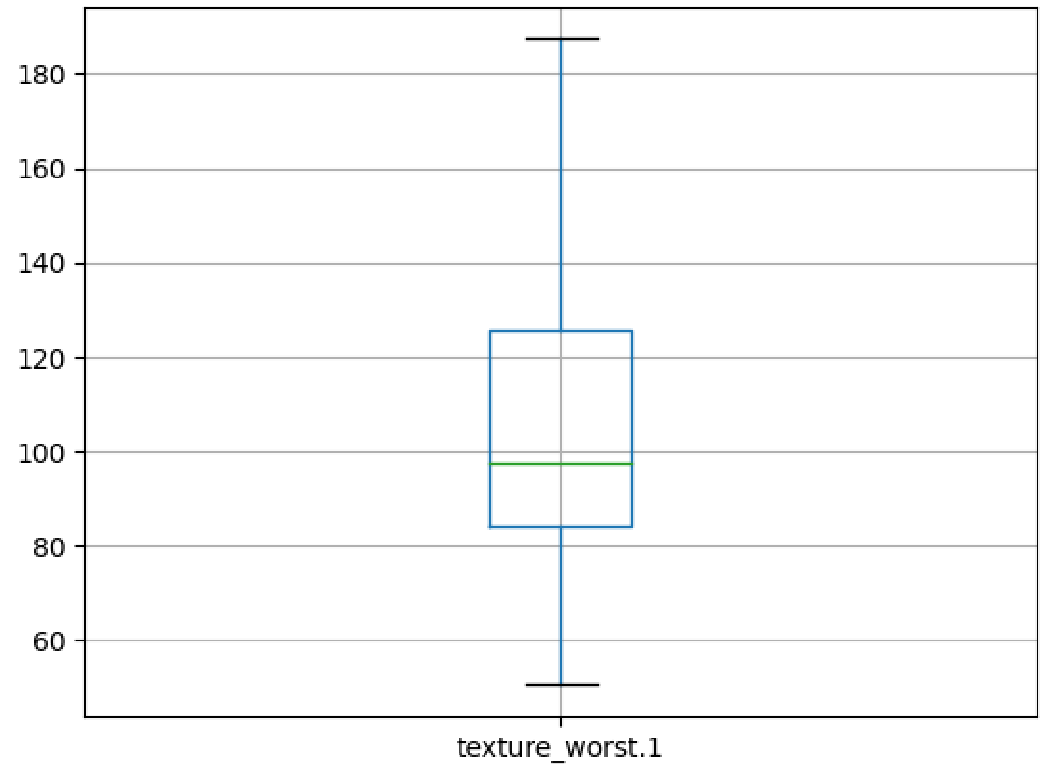

0.025
0.020
0.015
0.010
0.005
concave_points_se

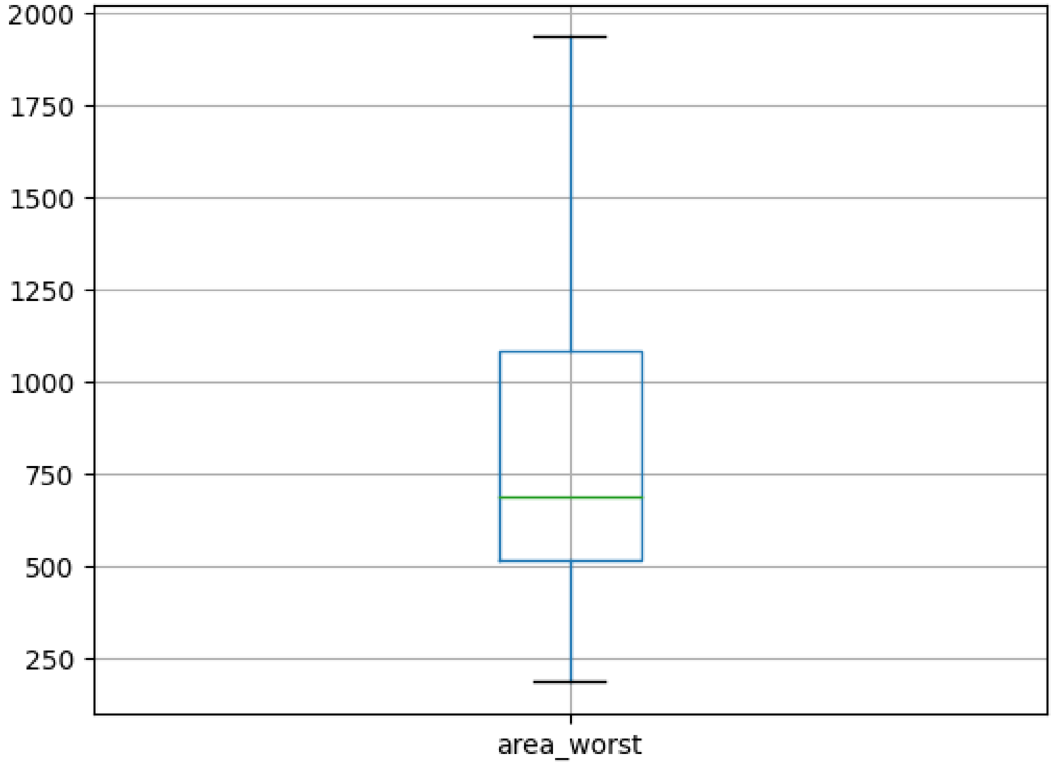

0.035
0.030
0.025
0.020
0.015
0.010
symmetry_se

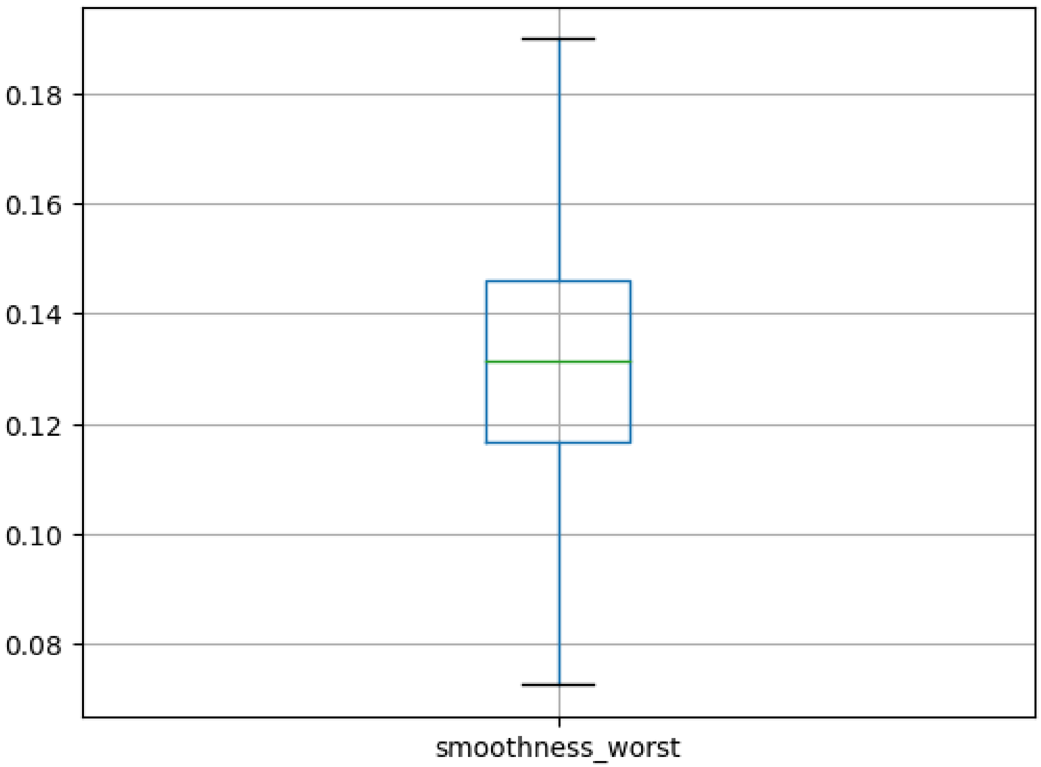

0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
fractal_dimension_se

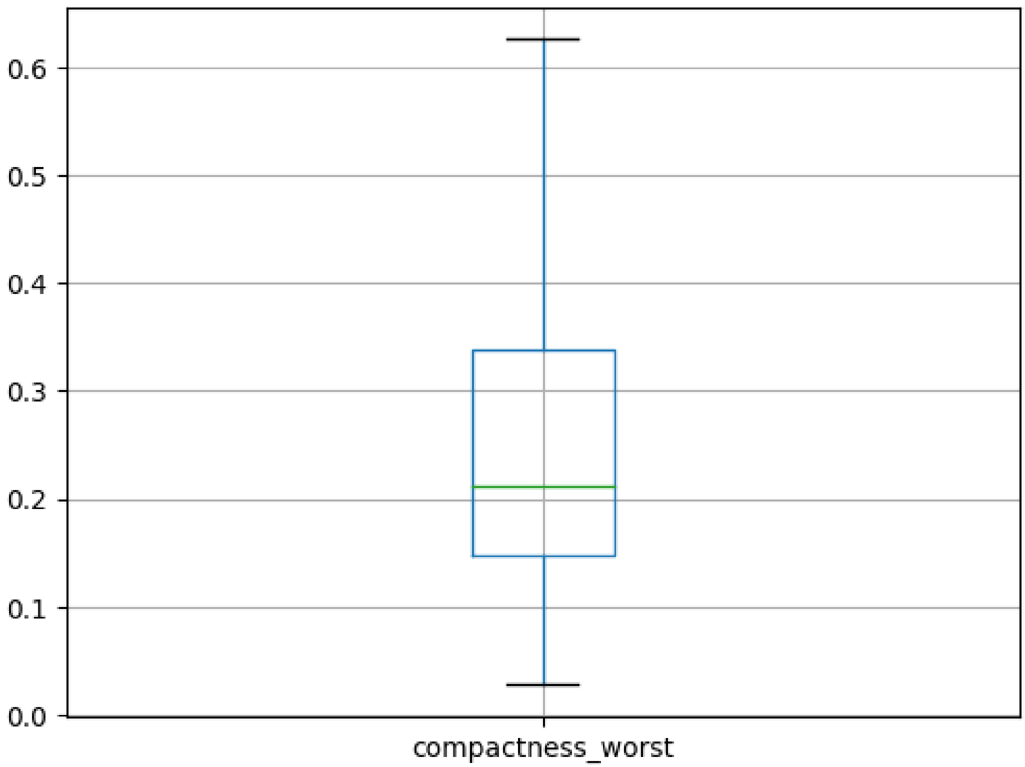

27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
radius_worst

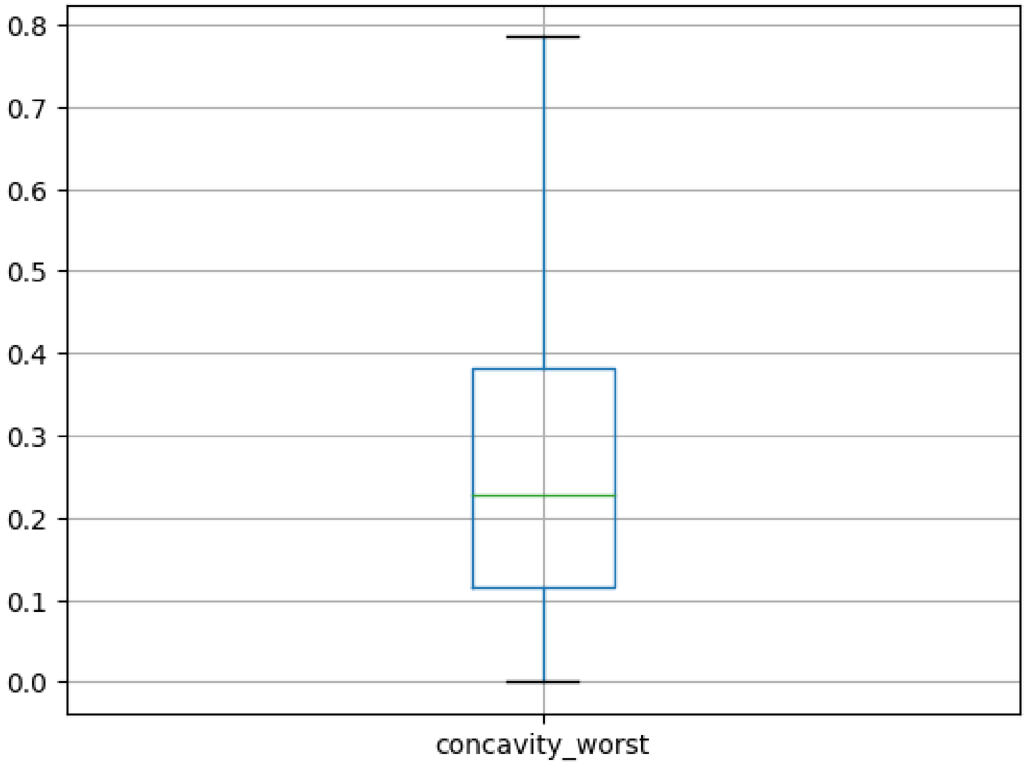

40
35
30
25
20
15
texture_worst

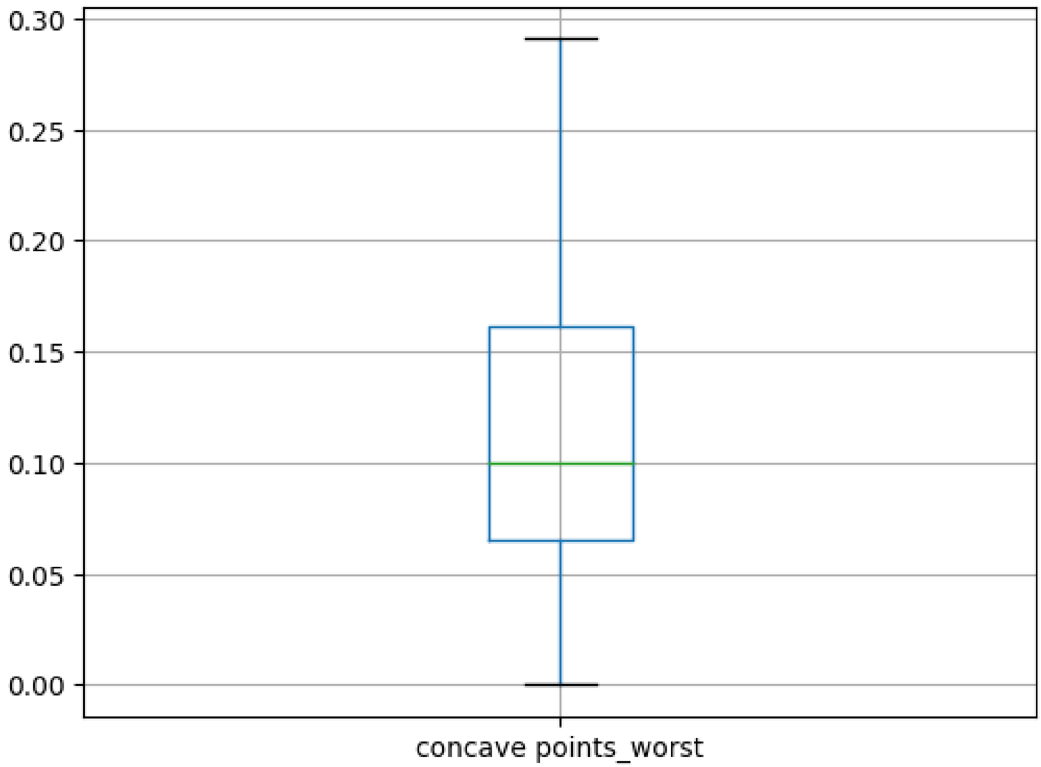

180
160
140
120
100
80
60
texture_worst.1

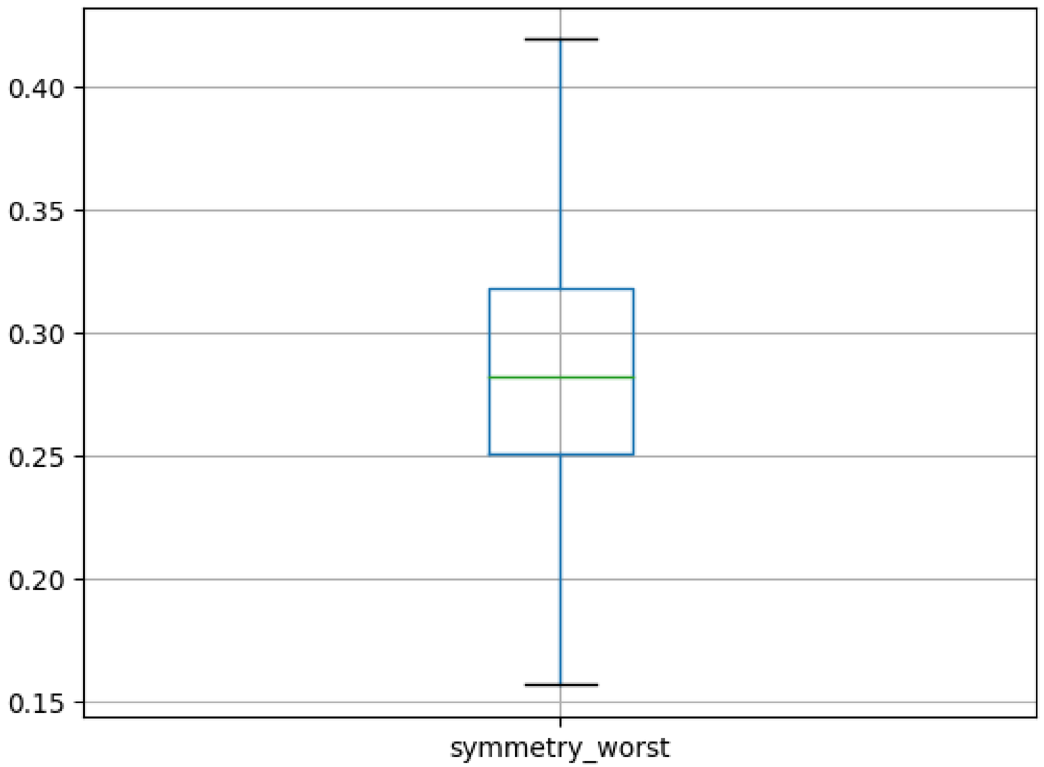

2000
1750
1500
1250
1000
750
500
250
area_worst

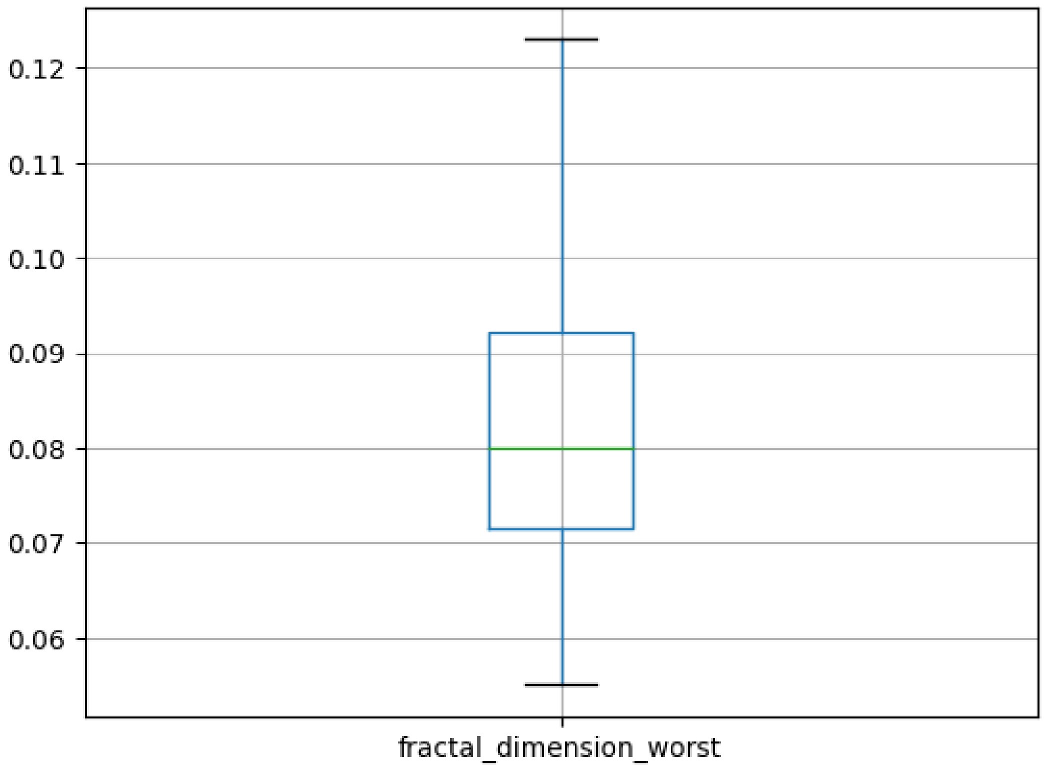

0.18
0.16
0.14
0.12
0.10
0.08
smoothness_worst


0.6
0.5
0.4
0.3
0.2
0.1
0.0
compactness_worst


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
concavity_worst


0.30
0.25
0.20
0.15
0.10
0.05
0.00
concave points_worst


0.40
0.35
0.30
0.25
0.20
0.15
symmetry_worst

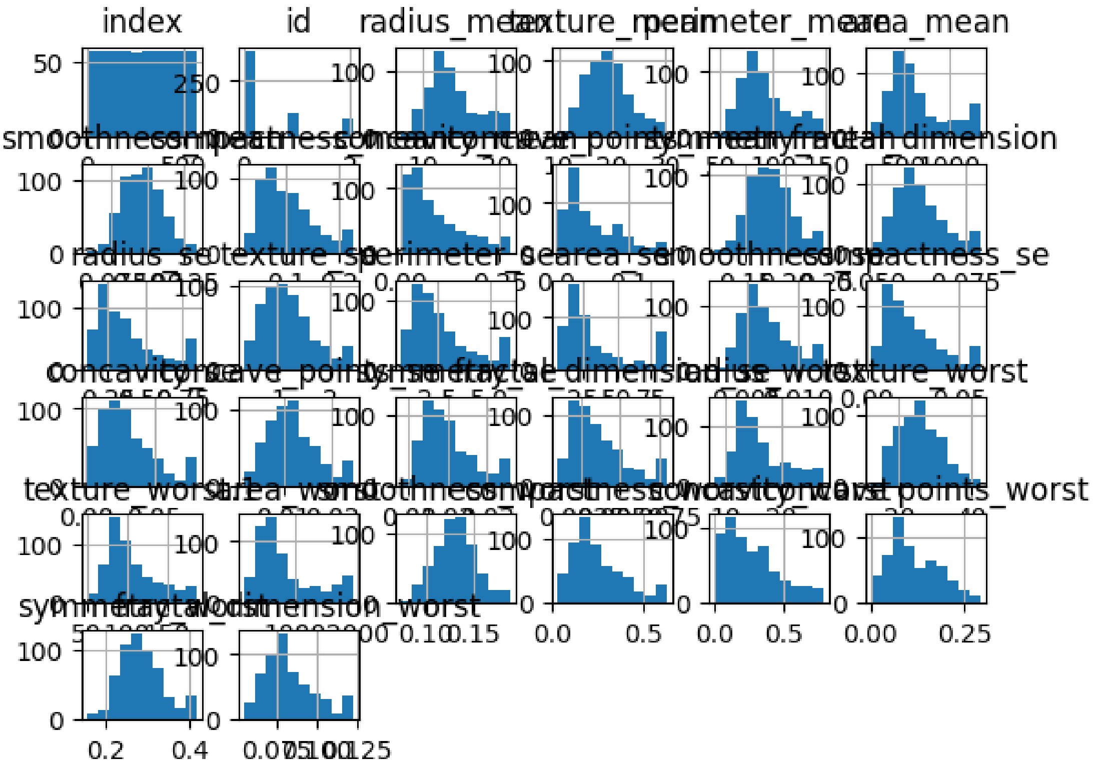

0.12
0.11
0.10
0.09
0.08
0.07
0.06
fractal_dimension_worst

[21]:

```
cancer_data.hist()
```

[21]:

```
array([[<Axes: title={'center': 'index'}>,
        <Axes: title={'center': 'id'}>,
        <Axes: title={'center': 'radius_mean'}>,
        <Axes: title={'center': 'texture_mean'}>,
        <Axes: title={'center': 'perimeter_mean'}>,
        <Axes: title={'center': 'area_mean'}>],
       [<Axes: title={'center': 'smoothness_mean'}>,
        <Axes: title={'center': 'compactness_mean'}>,
        <Axes: title={'center': 'concavity_mean'}>,
        <Axes: title={'center': 'concave_points_mean'}>,
        <Axes: title={'center': 'symmetry_mean'}>,
        <Axes: title={'center': 'fractal_dimension'}>],
       [<Axes: title={'center': 'radius_se'}>,
        <Axes: title={'center': 'texture_se'}>,
        <Axes: title={'center': 'perimeter_se'}>,
        <Axes: title={'center': 'area_se'}>,
        <Axes: title={'center': 'smoothness_se'}>,
        <Axes: title={'center': 'compactness_se'}>],
       [<Axes: title={'center': 'concavity_se'}>,
        <Axes: title={'center': 'concave_points_se'}>,
        <Axes: title={'center': 'symmetry_se'}>,
        <Axes: title={'center': 'fractal_dimension_se'}>,
        <Axes: title={'center': 'radius_worst'}>,
        <Axes: title={'center': 'texture_worst'}>],
       [<Axes: title={'center': 'texture_worst.1'}>,
        <Axes: title={'center': 'area_worst'}>,
        <Axes: title={'center': 'smoothness_worst'}>,
        <Axes: title={'center': 'compactness_worst'}>,
        <Axes: title={'center': 'concavity_worst'}>,
        <Axes: title={'center': 'concave points_worst'}>],
       [<Axes: title={'center': 'symmetry_worst'}>,
        <Axes: title={'center': 'fractal_dimension_worst'}>, <Axes: >,
        <Axes: >, <Axes: >, <Axes: >]], dtype=object)
```



[22]:

```
cancer_data.skew()
```

```
<ipython-input-22-a3ad186019e1>:1: FutureWarning: The default value of
numeric_only in DataFrame.skew is deprecated. In a future version, it will
default to False. In addition, specifying 'numeric_only=None' is deprecated.
Select only valid columns or specify the value of numeric_only to silence this
warning.
  cancer_data.skew()
```

[22]:

```
index                      0.000000
id                         1.418571
radius_mean                0.655953
texture_mean               0.449700
perimeter_mean             0.701081
area_mean                  0.922884
smoothness_mean            0.257712
compactness_mean           0.826755
concavity_mean             1.030074
concave_points_mean        1.047545
symmetry_mean              0.403661
fractal_dimension          0.682430
radius_se                  1.025031
```

```
texture_se                 0.740987
perimeter_se               1.034389
area_se                    1.130940
smoothness_se              0.780923
compactness_se             0.990285
concavity_se               0.950588
concave_points_se          0.682343
symmetry_se                0.869297
fractal_dimension_se       0.979344
radius_worst               0.849779
texture_worst              0.386858
texture_worst.1            0.874870
area_worst                 1.048970
smoothness_worst           0.247199
compactness_worst          0.919246
concavity_worst            0.808771
concave points_worst       0.492616
symmetry_worst             0.521772
fractal_dimension_worst    0.831581
dtype: float64
```

[23]:

```
cancer_data['diagnosis'].value_counts()
```

[23]:

```
B    357
M    212
Name: diagnosis, dtype: int64
```

[24]:

```
from sklearn.preprocessing import LabelEncoder
# loading the label Encoder function
label_encode=LabelEncoder()
labels=label_encode.fit_transform(cancer_data.diagnosis)
# appending labels to the dataframe
cancer_data['target']=labels
```

[25]:

```
cancer_data['target'].value_counts()
```

[25]:

```
0    357
1    212
Name: target, dtype: int64
```

[26]:

```
cancer_data
```

[26]:

```
     index          id  radius_mean  texture_mean  perimeter_mean  area_mean  \
0        1    842302.0        17.99         10.38          122.80     1001.0
1        2    842517.0        20.57         17.77          132.90     1326.0
2        3  20728995.5        19.69         21.25          130.00     1203.0
3        4  20728995.5        11.42         20.38           77.58      386.1
```

```
4        5  20728995.5        20.29         14.34          135.10     1297.0
..     …         …          …           …             …        …
564    565    926424.0        21.56         22.39          142.00     1326.3
565    566    926682.0        20.13         28.25          131.20     1261.0
566    567    926954.0        16.60         28.08          108.30      858.1
567    568    927241.0        20.60         29.33          140.10     1265.0
568    569     92751.0         7.76         24.54           47.92      181.0

     smoothness_mean  compactness_mean  concavity_mean  concave_points_mean  \
0           0.118400           0.22862        0.275355             0.147100
1           0.084740           0.07864        0.086900             0.070170
2           0.109600           0.15990        0.197400             0.127900
3           0.133695           0.22862        0.241400             0.105200
4           0.100300           0.13280        0.198000             0.104300
..                 …               …             …                  …
564         0.111000           0.11590        0.243900             0.138900
565         0.097800           0.10340        0.144000             0.097910
566         0.084550           0.10230        0.092510             0.053020
567         0.117800           0.22862        0.275355             0.151970
568         0.057975           0.04362        0.064905             0.034605

     …  compactness_worst  concavity_worst  concave points_worst  \
0    …            0.62620           0.7119                0.2654
1    …            0.18660           0.2416                0.1860
2    …            0.42450           0.4504                0.2430
3    …            0.62620           0.6869                0.2575
4    …            0.20500           0.4000                0.1625
..   …                …              …                   …
564  …            0.21130           0.4107                0.2216
565  …            0.19220           0.3215                0.1628
566  …            0.30940           0.3403                0.1418
567  …            0.62620           0.7855                0.2650
568  …            0.06444           0.0000                0.0000

     symmetry_worst  fractal_dimension_worst  N Stage  6th Stage  \
0           0.41915                  0.11890       N1        IIA
1           0.27500                  0.08902       N2       IIIA
2           0.36130                  0.08758       N3       IIIC
3           0.41915                  0.12301       N1        IIA
4           0.23640                  0.07678       N1        IIB
..              …                      …      …        …
564         0.20600                  0.07115       N2       IIIA
565         0.25720                  0.06637       N2       IIIA
566         0.22180                  0.07820       N1        IIA
567         0.40870                  0.12301       N1        IIA
568         0.28710                  0.07039       N3       IIIC
```

```
                differentiate  diagnosis  target
0         Poorly differentiated          M       1
1     Moderately differentiated          M       1
2     Moderately differentiated          M       1
3         Poorly differentiated          M       1
4         Poorly differentiated          M       1
..                          …         …      …
564   Moderately differentiated          M       1
565       Poorly differentiated          M       1
566         Well differentiated          M       1
567         Well differentiated          M       1
568         Well differentiated          B       0

[569 rows x 37 columns]
```

[28]:

```python
# data standardization
from sklearn.preprocessing import  StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

scaler=StandardScaler()
```

[ ]:

```

```

[ ]:

```
     index          id  radius_mean  texture_mean  perimeter_mean  area_mean  \
0        1    842302.0        17.99         10.38          122.80     1001.0
1        2    842517.0        20.57         17.77          132.90     1326.0
2        3  20728995.5        19.69         21.25          130.00     1203.0
3        4  20728995.5        11.42         20.38           77.58      386.1
4        5  20728995.5        20.29         14.34          135.10     1297.0
..     …           …          …            …             …          …
564    565    926424.0        21.56         22.39          142.00     1326.3
565    566    926682.0        20.13         28.25          131.20     1261.0
566    567    926954.0        16.60         28.08          108.30      858.1
567    568    927241.0        20.60         29.33          140.10     1265.0
568    569     92751.0         7.76         24.54           47.92      181.0

     smoothness_mean  compactness_mean  concavity_mean  concave_points_mean  \
0           0.118400           0.22862        0.275355             0.147100
1           0.084740           0.07864        0.086900             0.070170
2           0.109600           0.15990        0.197400             0.127900
3           0.133695           0.22862        0.241400             0.105200
4           0.100300           0.13280        0.198000             0.104300
..               …               …             …                   …
564         0.111000           0.11590        0.243900             0.138900
565         0.097800           0.10340        0.144000             0.097910
566         0.084550           0.10230        0.092510             0.053020
```

```
567         0.117800           0.22862        0.275355             0.151970
568         0.057975           0.04362        0.064905             0.034605

     …  smoothness_worst  compactness_worst  concavity_worst  \
0    …         0.16220            0.62620           0.7119
1    …         0.12380            0.18660           0.2416
2    …         0.14440            0.42450           0.4504
3    …         0.19010            0.62620           0.6869
4    …         0.13740            0.20500           0.4000
..   …               …                …              …
564  …         0.14100            0.21130           0.4107
565  …         0.11660            0.19220           0.3215
566  …         0.11390            0.30940           0.3403
567  …         0.16500            0.62620           0.7855
568  …         0.08996            0.06444           0.0000

     concave points_worst  symmetry_worst  fractal_dimension_worst  N Stage  \
0                  0.2654         0.41915                  0.11890       N1
1                  0.1860         0.27500                  0.08902       N2
2                  0.2430         0.36130                  0.08758       N3
3                  0.2575         0.41915                  0.12301       N1
4                  0.1625         0.23640                  0.07678       N1
..                    …             …                      …      …
564                0.2216         0.20600                  0.07115       N2
565                0.1628         0.25720                  0.06637       N2
566                0.1418         0.22180                  0.07820       N1
567                0.2650         0.40870                  0.12301       N1
568                0.0000         0.28710                  0.07039       N3

     6th Stage              differentiate  diagnosis
0          IIA      Poorly differentiated          M
1         IIIA  Moderately differentiated          M
2         IIIC  Moderately differentiated          M
3          IIA      Poorly differentiated          M
4          IIB      Poorly differentiated          M
..         …                        …        …
564       IIIA  Moderately differentiated          M
565       IIIA      Poorly differentiated          M
566        IIA        Well differentiated          M
567        IIA        Well differentiated          M
568       IIIC        Well differentiated          B

[569 rows x 36 columns]
```

[32]:

```
from sklearn.linear_model import LogisticRegression
# classifier=Logistic_Regression(random_state=16)
# classifier.fit(x_train,y_train)
```

[36]:

```
model=LogisticRegression()
```

[43]:

```
cancer_data.groupby('target').mean()
```

```
<ipython-input-43-d70956e990b4>:1: FutureWarning: The default value of
numeric_only in DataFrameGroupBy.mean is deprecated. In a future version,
numeric_only will default to False. Either specify numeric_only or select only
columns which should be valid for the function.
  cancer_data.groupby('target').mean()
```

[43]:

```
             index            id  radius_mean  texture_mean  perimeter_mean  \
target
0        320.89916  4.746260e+06    12.146524     17.903445       78.075406
1        224.54717  5.699931e+06    17.290047     21.530259      114.224033

         area_mean  smoothness_mean  compactness_mean  concavity_mean  \
target
0       462.790196         0.092400          0.080085        0.048081
1       937.784434         0.102776          0.142184        0.155316

        concave_points_mean  …  radius_worst  texture_worst  \
target                       …
0                  0.027032  …     13.379801      23.515070
1                  0.086432  …     20.905849      29.240991

        texture_worst.1   area_worst  smoothness_worst  compactness_worst  \
target
0             87.005938   558.899440          0.124934           0.182993
1            139.878467  1339.954953          0.144462           0.363006

        concavity_worst  concave points_worst  symmetry_worst  \
target
0              0.164808              0.074444        0.270236
1              0.443757              0.182237        0.316884

        fractal_dimension_worst
target
0                      0.079295
1                      0.090159

[2 rows x 32 columns]
```

[41]:

```
cancer_data
```

[41]:

```
      index        id  radius_mean  texture_mean  perimeter_mean  area_mean  \
0         1  842302.0        17.99         10.38          122.80     1001.0
1         2  842517.0        20.57         17.77          132.90     1326.0
```

```
2        3  20728995.5        19.69          21.25            130.00     1203.0
3        4  20728995.5        11.42          20.38             77.58      386.1
4        5  20728995.5        20.29          14.34            135.10     1297.0
..     …         …          …           …             …        …
564    565    926424.0        21.56          22.39            142.00     1326.3
565    566    926682.0        20.13          28.25            131.20     1261.0
566    567    926954.0        16.60          28.08            108.30      858.1
567    568    927241.0        20.60          29.33            140.10     1265.0
568    569     92751.0         7.76          24.54             47.92      181.0

     smoothness_mean  compactness_mean  concavity_mean  concave_points_mean  \
0           0.118400           0.22862        0.275355            0.147100
1           0.084740           0.07864        0.086900            0.070170
2           0.109600           0.15990        0.197400            0.127900
3           0.133695           0.22862        0.241400            0.105200
4           0.100300           0.13280        0.198000            0.104300
..               …               …             …                  …
564         0.111000           0.11590        0.243900            0.138900
565         0.097800           0.10340        0.144000            0.097910
566         0.084550           0.10230        0.092510            0.053020
567         0.117800           0.22862        0.275355            0.151970
568         0.057975           0.04362        0.064905            0.034605

     …  compactness_worst  concavity_worst  concave points_worst  \
0    …            0.62620           0.7119                0.2654
1    …            0.18660           0.2416                0.1860
2    …            0.42450           0.4504                0.2430
3    …            0.62620           0.6869                0.2575
4    …            0.20500           0.4000                0.1625
..   …                …              …                   …
564  …            0.21130           0.4107                0.2216
565  …            0.19220           0.3215                0.1628
566  …            0.30940           0.3403                0.1418
567  …            0.62620           0.7855                0.2650
568  …            0.06444           0.0000                0.0000

     symmetry_worst  fractal_dimension_worst  N Stage  6th Stage  \
0           0.41915                  0.11890       N1        IIA
1           0.27500                  0.08902       N2       IIIA
2           0.36130                  0.08758       N3       IIIC
3           0.41915                  0.12301       N1        IIA
4           0.23640                  0.07678       N1        IIB
..              …                      …      …        …
564         0.20600                  0.07115       N2       IIIA
565         0.25720                  0.06637       N2       IIIA
566         0.22180                  0.07820       N1        IIA
567         0.40870                  0.12301       N1        IIA
```

```
568          0.28710                  0.07039       N3       IIIC

                 differentiate diagnosis  target
0        Poorly differentiated         M       1
1    Moderately differentiated         M       1
2    Moderately differentiated         M       1
3        Poorly differentiated         M       1
4        Poorly differentiated         M       1
..                         …        …      …
564  Moderately differentiated         M       1
565      Poorly differentiated         M       1
566        Well differentiated         M       1
567        Well differentiated         M       1
568        Well differentiated         B       0

[569 rows x 37 columns]
```

[53]:
```
new_cancer_data=cancer_data.drop(columns='diagnosis',axis=1)
```

[49]:
```
string_columns = cancer_data.select_dtypes(include=['object']).columns
```

[50]:
```
string_columns
```

[50]:
```
Index(['N Stage', '6th Stage', 'differentiate', 'diagnosis'], dtype='object')
```

[51]:
```
cancer_data
```

[51]:
```
     index          id  radius_mean  texture_mean  perimeter_mean  area_mean  \
0        1    842302.0        17.99         10.38          122.80     1001.0
1        2    842517.0        20.57         17.77          132.90     1326.0
2        3  20728995.5        19.69         21.25          130.00     1203.0
3        4  20728995.5        11.42         20.38           77.58      386.1
4        5  20728995.5        20.29         14.34          135.10     1297.0
..     …         …          …            …             …         …
564    565    926424.0        21.56         22.39          142.00     1326.3
565    566    926682.0        20.13         28.25          131.20     1261.0
566    567    926954.0        16.60         28.08          108.30      858.1
567    568    927241.0        20.60         29.33          140.10     1265.0
568    569     92751.0         7.76         24.54           47.92      181.0

     smoothness_mean  compactness_mean  concavity_mean  concave_points_mean  \
0           0.118400           0.22862        0.275355             0.147100
1           0.084740           0.07864        0.086900             0.070170
2           0.109600           0.15990        0.197400             0.127900
3           0.133695           0.22862        0.241400             0.105200
4           0.100300           0.13280        0.198000             0.104300
..               …               …             …                   …
```

```
564         0.111000           0.11590        0.243900             0.138900
565         0.097800           0.10340        0.144000             0.097910
566         0.084550           0.10230        0.092510             0.053020
567         0.117800           0.22862        0.275355             0.151970
568         0.057975           0.04362        0.064905             0.034605

     …  compactness_worst  concavity_worst  concave points_worst  \
0    …            0.62620           0.7119                0.2654
1    …            0.18660           0.2416                0.1860
2    …            0.42450           0.4504                0.2430
3    …            0.62620           0.6869                0.2575
4    …            0.20500           0.4000                0.1625
..   …                  …                …                     …
564  …            0.21130           0.4107                0.2216
565  …            0.19220           0.3215                0.1628
566  …            0.30940           0.3403                0.1418
567  …            0.62620           0.7855                0.2650
568  …            0.06444           0.0000                0.0000

     symmetry_worst  fractal_dimension_worst  N Stage  6th Stage  \
0           0.41915                  0.11890       N1        IIA
1           0.27500                  0.08902       N2       IIIA
2           0.36130                  0.08758       N3       IIIC
3           0.41915                  0.12301       N1        IIA
4           0.23640                  0.07678       N1        IIB
..                …                        …      …        …
564         0.20600                  0.07115       N2       IIIA
565         0.25720                  0.06637       N2       IIIA
566         0.22180                  0.07820       N1        IIA
567         0.40870                  0.12301       N1        IIA
568         0.28710                  0.07039       N3       IIIC

                 differentiate  diagnosis  target
0        Poorly differentiated          M       1
1    Moderately differentiated          M       1
2    Moderately differentiated          M       1
3        Poorly differentiated          M       1
4        Poorly differentiated          M       1
..                          …         …      …
564  Moderately differentiated          M       1
565      Poorly differentiated          M       1
566        Well differentiated          M       1
567        Well differentiated          M       1
568        Well differentiated          B       0

[569 rows x 37 columns]
```

```
[52]: cancer_data.shape
```

```
[52]: (569, 37)
```

```
[54]: new_cancer_data
```

```
[54]:      index          id  radius_mean  texture_mean  perimeter_mean  area_mean  \
      0        1    842302.0        17.99         10.38          122.80     1001.0
      1        2    842517.0        20.57         17.77          132.90     1326.0
      2        3  20728995.5        19.69         21.25          130.00     1203.0
      3        4  20728995.5        11.42         20.38           77.58      386.1
      4        5  20728995.5        20.29         14.34          135.10     1297.0
      ..     …           …            …            …              …          …
      564    565    926424.0        21.56         22.39          142.00     1326.3
      565    566    926682.0        20.13         28.25          131.20     1261.0
      566    567    926954.0        16.60         28.08          108.30      858.1
      567    568    927241.0        20.60         29.33          140.10     1265.0
      568    569     92751.0         7.76         24.54           47.92      181.0

           smoothness_mean  compactness_mean  concavity_mean  concave_points_mean  \
      0           0.118400           0.22862        0.275355             0.147100
      1           0.084740           0.07864        0.086900             0.070170
      2           0.109600           0.15990        0.197400             0.127900
      3           0.133695           0.22862        0.241400             0.105200
      4           0.100300           0.13280        0.198000             0.104300
      ..               …                 …              …                    …
      564         0.111000           0.11590        0.243900             0.138900
      565         0.097800           0.10340        0.144000             0.097910
      566         0.084550           0.10230        0.092510             0.053020
      567         0.117800           0.22862        0.275355             0.151970
      568         0.057975           0.04362        0.064905             0.034605

           …  smoothness_worst  compactness_worst  concavity_worst  \
      0    …           0.16220            0.62620           0.7119
      1    …           0.12380            0.18660           0.2416
      2    …           0.14440            0.42450           0.4504
      3    …           0.19010            0.62620           0.6869
      4    …           0.13740            0.20500           0.4000
      ..   …               …                  …                …
      564  …           0.14100            0.21130           0.4107
      565  …           0.11660            0.19220           0.3215
      566  …           0.11390            0.30940           0.3403
      567  …           0.16500            0.62620           0.7855
      568  …           0.08996            0.06444           0.0000

           concave points_worst  symmetry_worst  fractal_dimension_worst  N Stage  \
      0                  0.2654         0.41915                  0.11890       N1
```

```
1                  0.1860         0.27500                  0.08902       N2
2                  0.2430         0.36130                  0.08758       N3
3                  0.2575         0.41915                  0.12301       N1
4                  0.1625         0.23640                  0.07678       N1
..                    …             …                      …      …
564                0.2216         0.20600                  0.07115       N2
565                0.1628         0.25720                  0.06637       N2
566                0.1418         0.22180                  0.07820       N1
567                0.2650         0.40870                  0.12301       N1
568                0.0000         0.28710                  0.07039       N3

     6th Stage              differentiate  target
0          IIA      Poorly differentiated       1
1         IIIA  Moderately differentiated       1
2         IIIC  Moderately differentiated       1
3          IIA      Poorly differentiated       1
4          IIB      Poorly differentiated       1
..         …                        …     …
564       IIIA  Moderately differentiated       1
565       IIIA      Poorly differentiated       1
566        IIA        Well differentiated       1
567        IIA        Well differentiated       1
568       IIIC        Well differentiated       0

[569 rows x 36 columns]
```

[56]:

```
string_columns = new_cancer_data.select_dtypes(include=['object']).columns
```

[57]:

```
string_columns
```

[57]:

```
Index(['N Stage', '6th Stage', 'differentiate'], dtype='object')
```

[58]:

```
# now convert all string column into integer value
new_cancer_data['6th Stage'].value_counts()
```

[58]:

```
IIA     194
IIB     150
IIIA    142
IIIC     77
IIIB      6
Name: 6th Stage, dtype: int64
```

[60]:

```
# separate data and labels in to apply logistic regression
features=new_cancer_data.drop(columns='target',axis=1);
tar=new_cancer_data['target']
```

[61]:
```
x_train,x_test, y_train,y_test=train_test_split(features,tar,test_size=0.2,␣
 ↪random_state=2)
```

[62]:
```
print(features.shape,x_train.shape, x_test.shape)
```

```
(569, 35) (455, 35) (114, 35)
```

[64]:
```
new_cancer_data=new_cancer_data.drop(columns='N Stage',axis=1)
new_cancer_data=new_cancer_data.drop(columns='6th Stage',axis=1)
new_cancer_data=new_cancer_data.drop(columns='differentiate',axis=1)
```

[66]:
```
new_cancer_data.shape
```

[66]:
```
(569, 33)
```

[67]:
```
# separate data and labels in to apply logistic regression
features=new_cancer_data.drop(columns='target',axis=1);
tar=new_cancer_data['target']
```

[68]:
```
x_train,x_test, y_train,y_test=train_test_split(features,tar,test_size=0.2,␣
 ↪random_state=2)
```

[69]:
```
print(features.shape,x_train.shape, x_test.shape)
```

```
(569, 32) (455, 32) (114, 32)
```

[70]:
```
model.fit(x_train,y_train)
```

[70]:
```
LogisticRegression()
```

[71]:
```
# model evaluation  apply on training data

x_train_pred=model.predict(x_train)
train_data_accuracy=accuracy_score(y_train,x_train_pred)
```

[72]:
```
print("Accuracy score of the train data ", train_data_accuracy)
```

```
Accuracy score of the train data  0.6329670329670329
```

[73]:
```
# model evaluation  apply on test data
x_test_pred=model.predict(x_test)
test_data_accuracy=accuracy_score(y_test,x_test_pred)
```

[74]:
```
print("Accuracy score of the test data ", test_data_accuracy)
```

```
Accuracy score of the test data  0.6052631578947368
```

[ ]:
```

```