

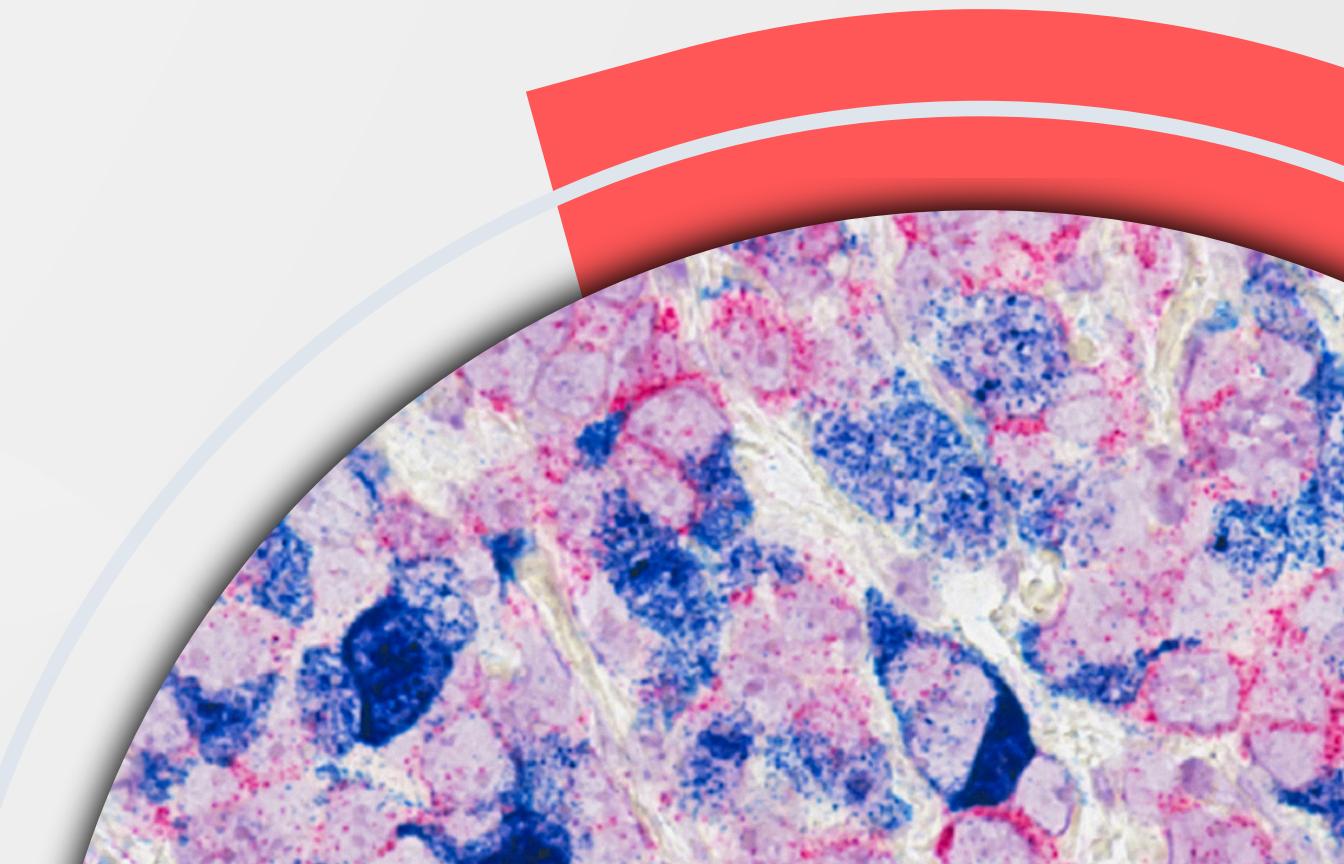
Refining Diagnostic Indicators:

Correlation Analysis of Overlapping Metrics in Breast Cancer Cell Nuclei

Prepared By

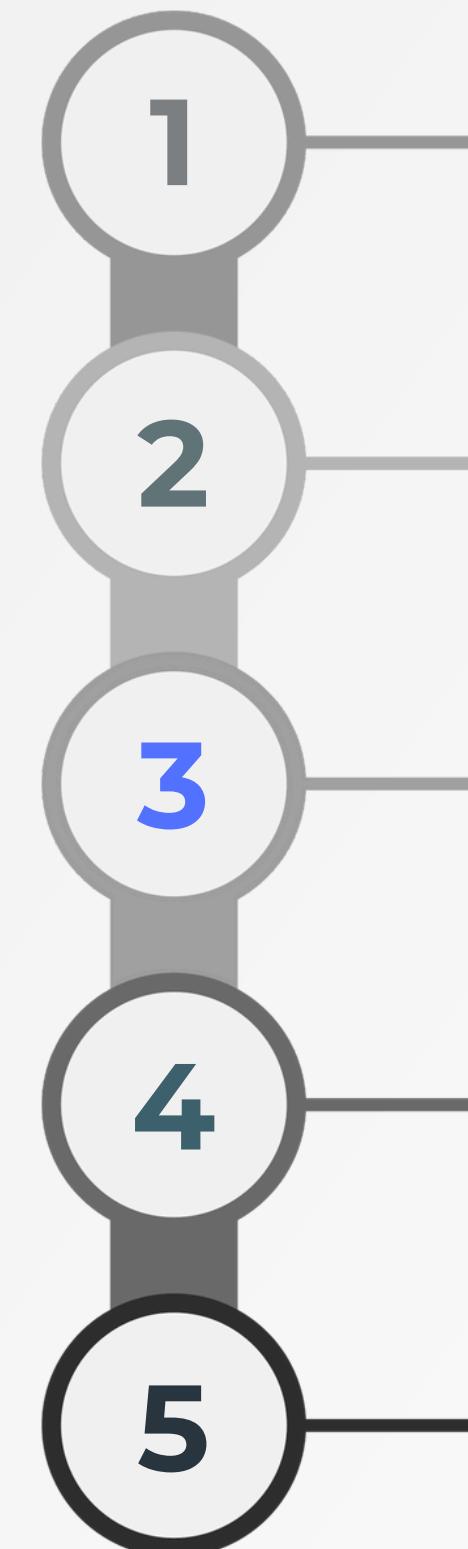
Vanuhi Davtyan

7th March, 2025



Purpose of analysis

This analysis is designed to identify the most influential nuclear features extracted from digitized FNA images that differentiate between malignant and benign breast masses. By focusing on overlapping metric ranges, the study seeks to pinpoint the key indicators most strongly correlated with the diagnosis, thereby enhancing the precision of computer-aided diagnostic methods.



Understand the data

Brief introduction to the analysed metrics and their values

Narrow down

To avoid redundancy reduction of the quantity of metrics to be analysed

Analyse the groups

Logical grouping and deeper dive into the correlation with diagnosis

Understand the correlation

Defining the correlation and association degrees with diagnosis

Draw conclusions

summary and overview the results of the performed analysis

Understanding the data

no missing values

569 records

30 numeric continuous columns

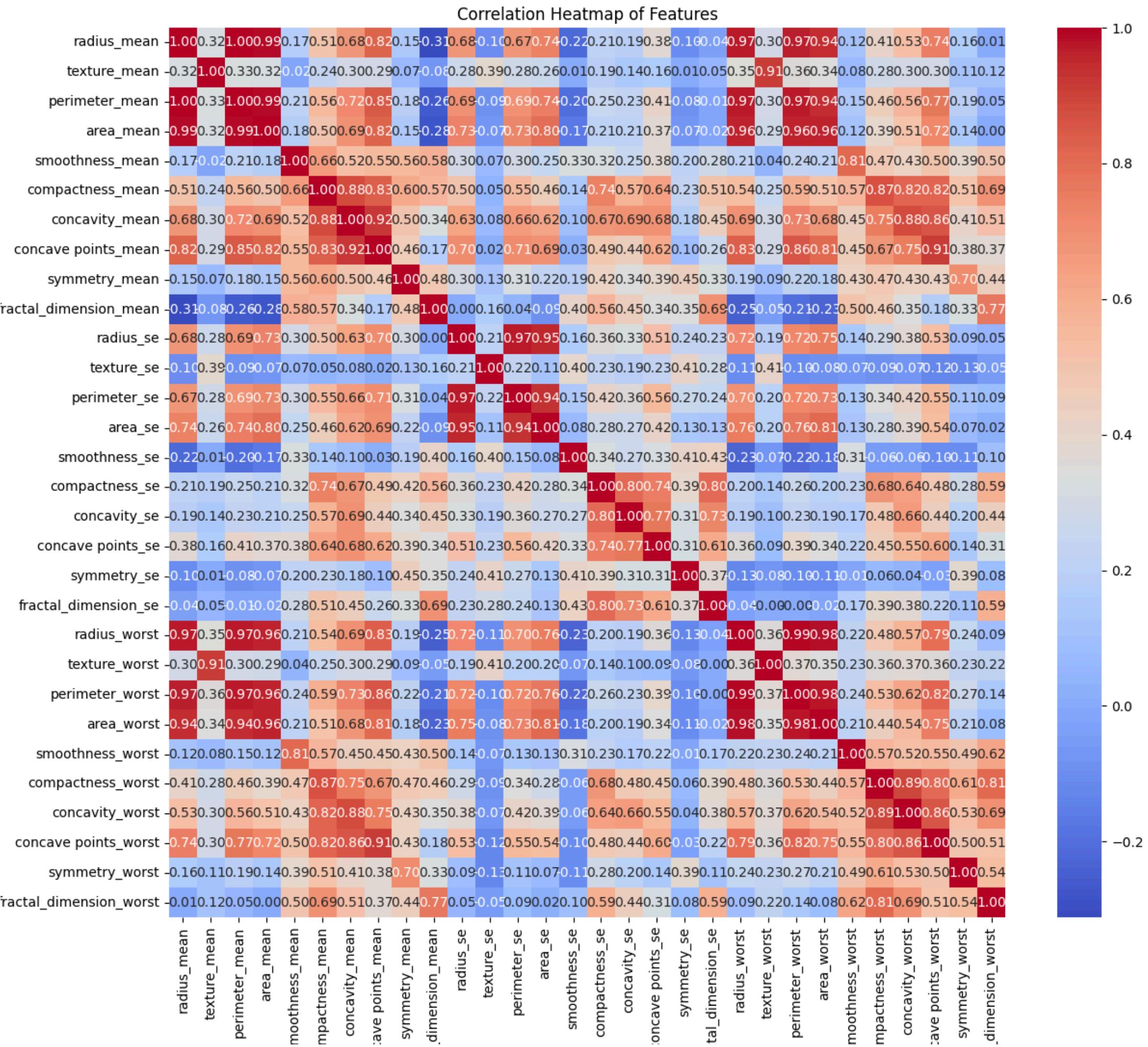
	Absolute Frequency	Relative Frequency
Benign	357	0.63
Malignant	212	0.37
Total	569	1.00



Correlation Matrix

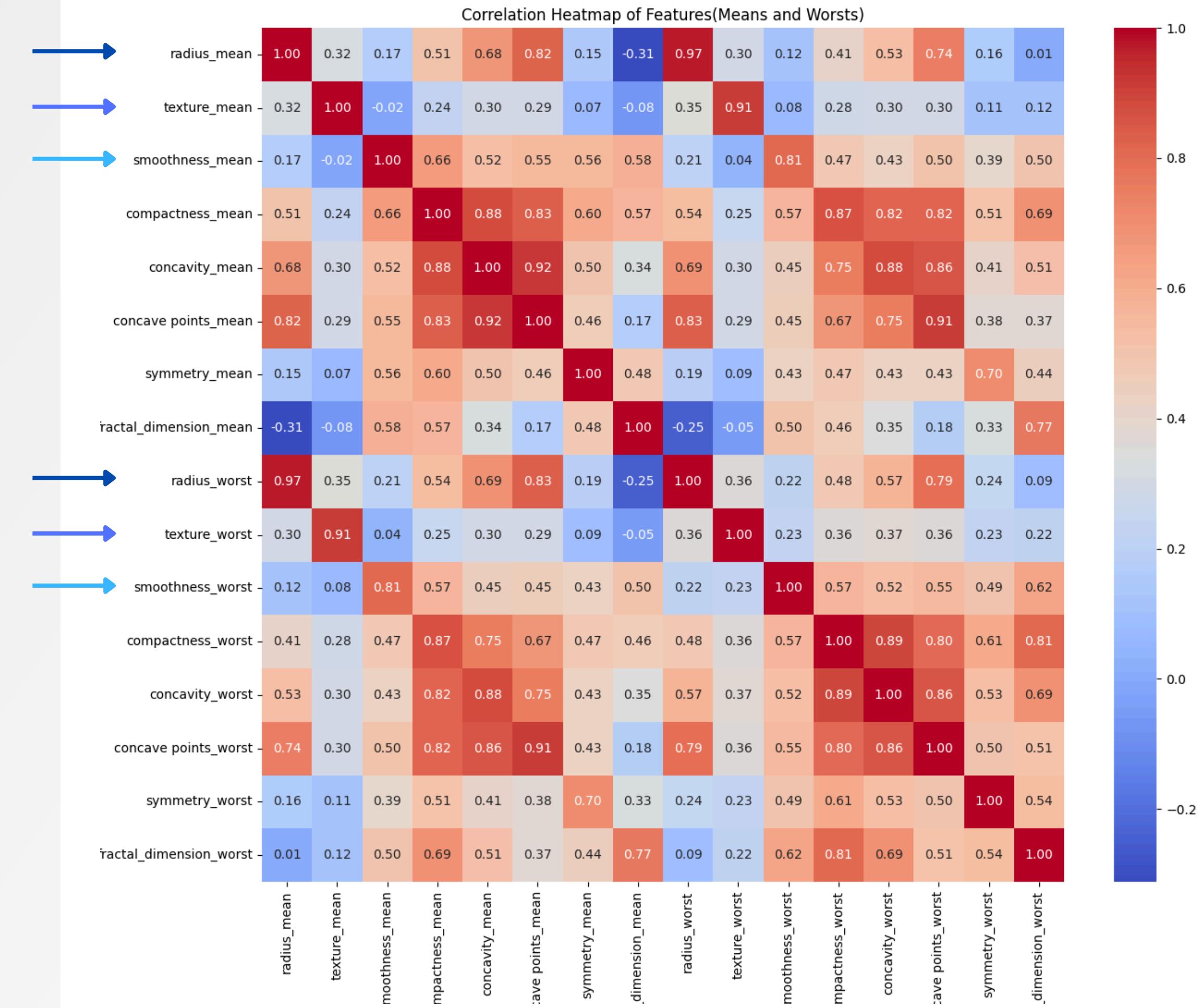
OBSERVATION

Radius, area and perimeter very high correlation values due to mathematical connection



Correlation Matrix

Idea: Dimensionality reduction - to reduce redundancy



radius

texture

smoothness

compactness

concavity

concave points

symmetry

fractal dimension



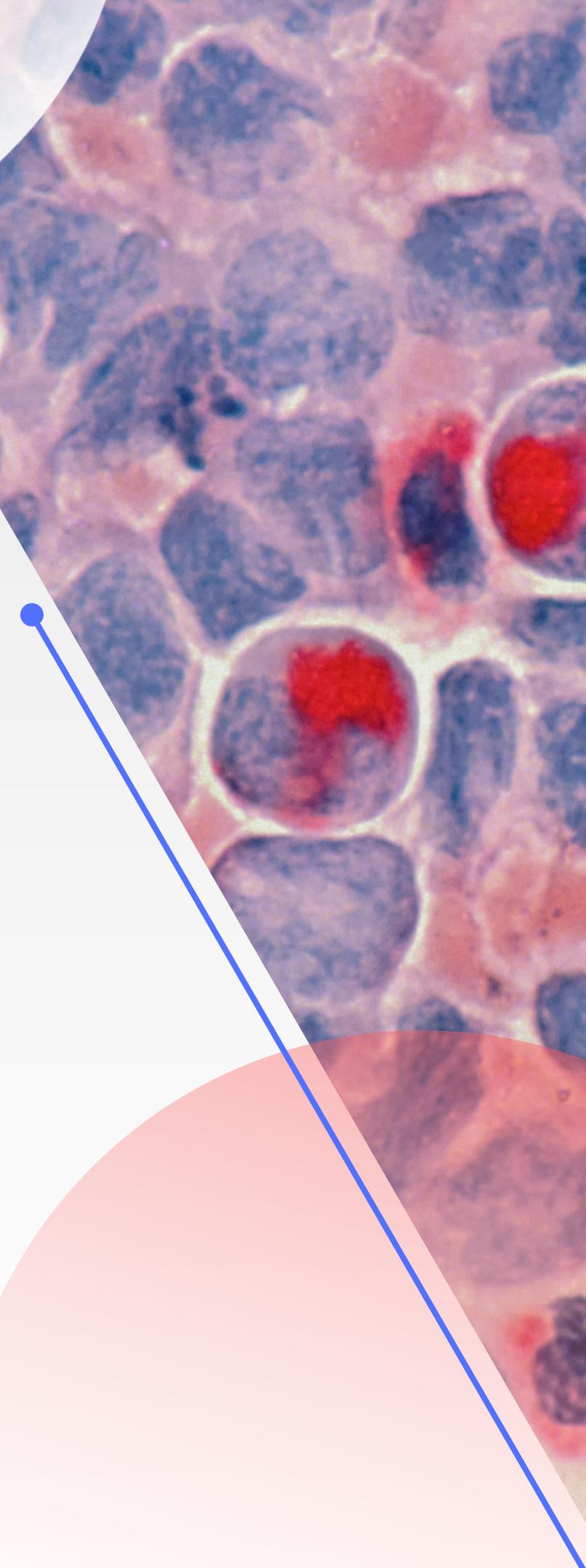
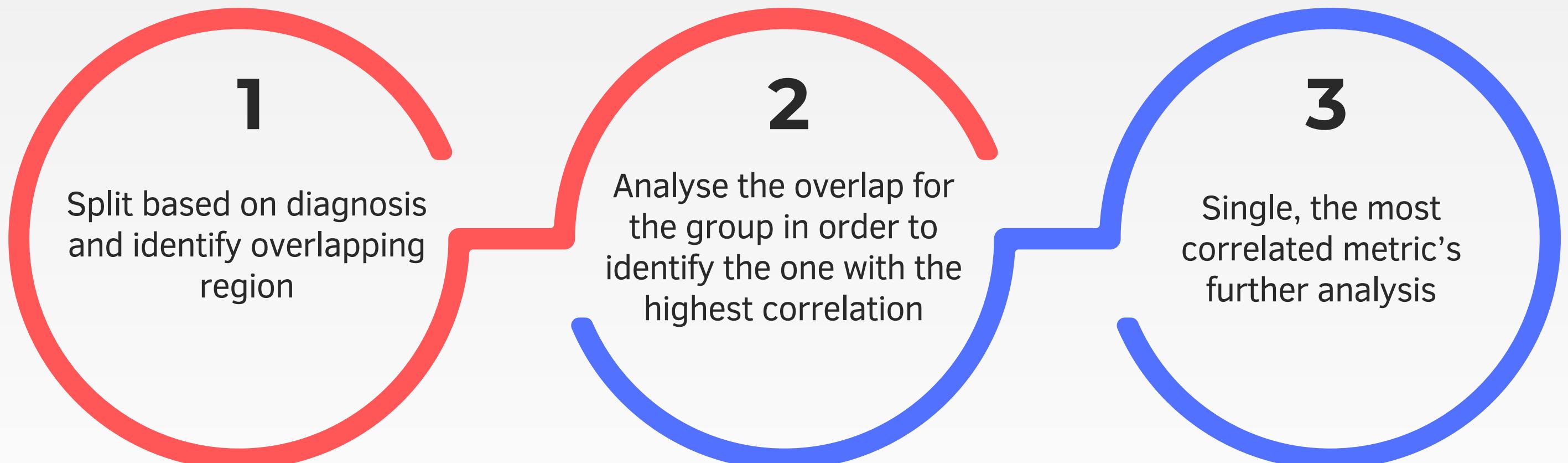
Analyse the Groups

Idea: Define a metric that would be independent of the absolute sizes and display the dynamic of change between mean and worst values

difference = worst - mean

Targeted group analysis

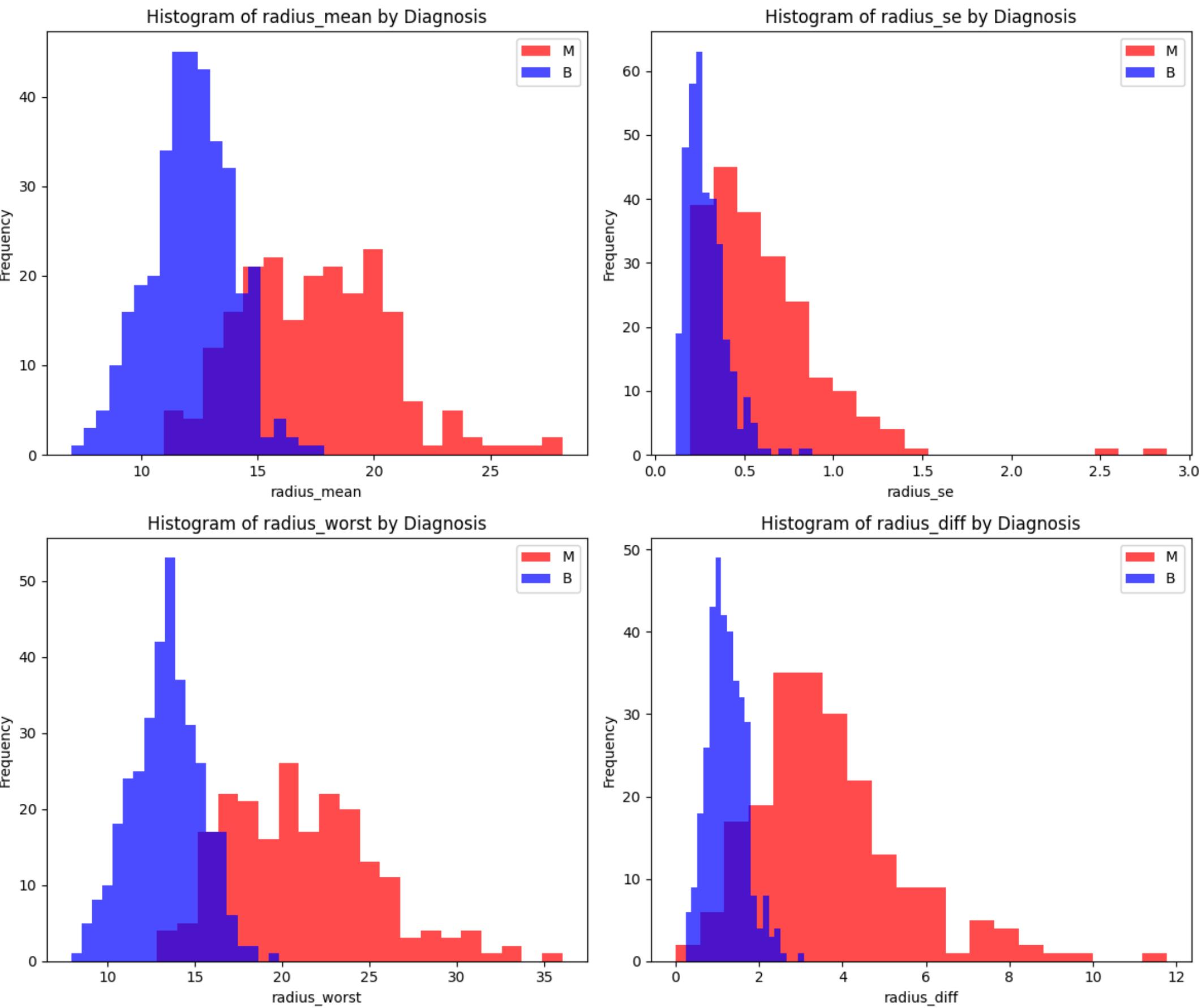
focus: radius



Overlaps

General observation:

Similarities in distribution of mean and worst, as well as of standart error and difference values for the radius

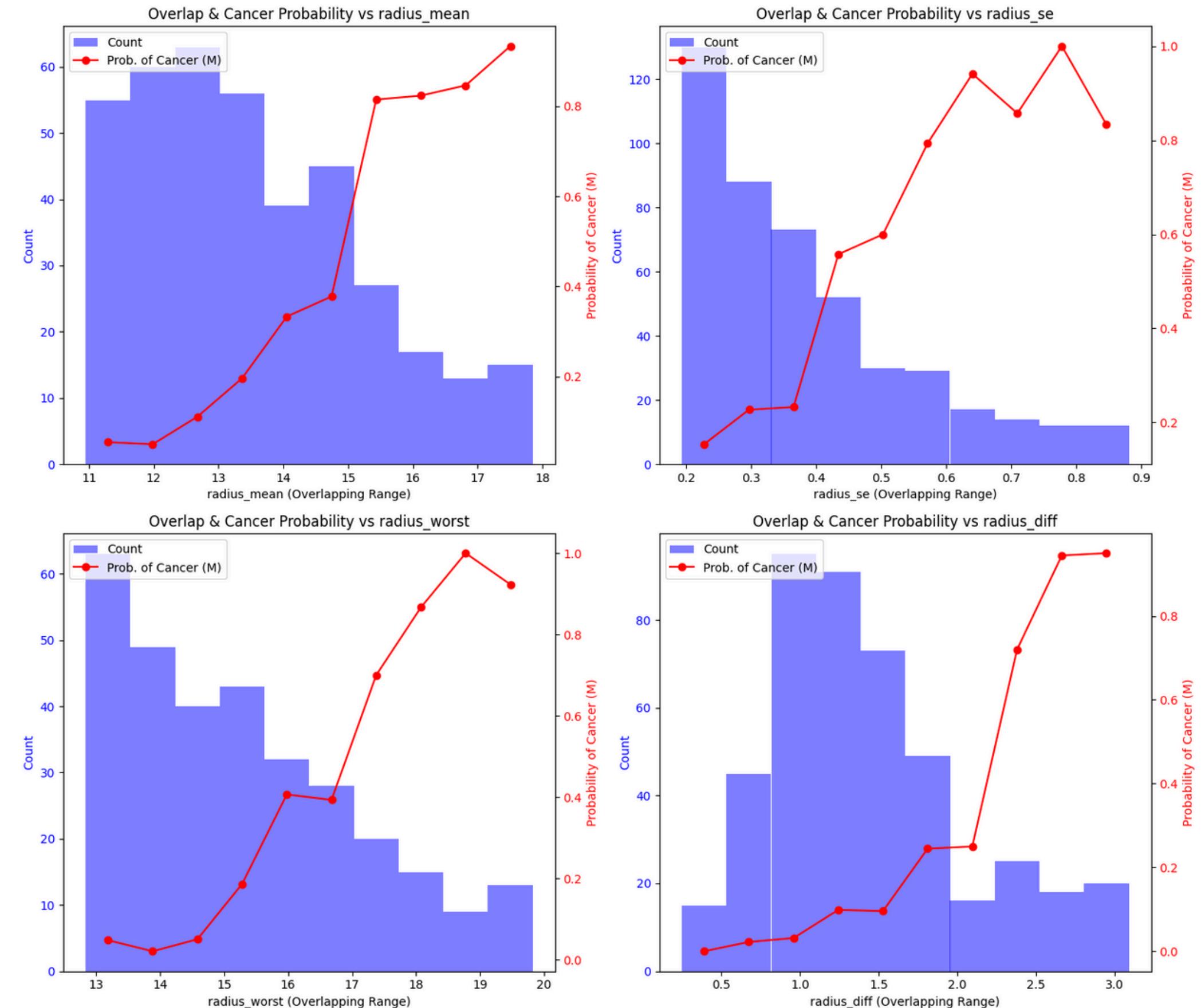


General observation:
Similarities in distribution of mean and
worst, as well as of standard error and
difference values for the radius

Overlaps

Why they matter?

Overlaps are the areas where it is
questionable whether the patient should be
diagnosed with benign or malignant.



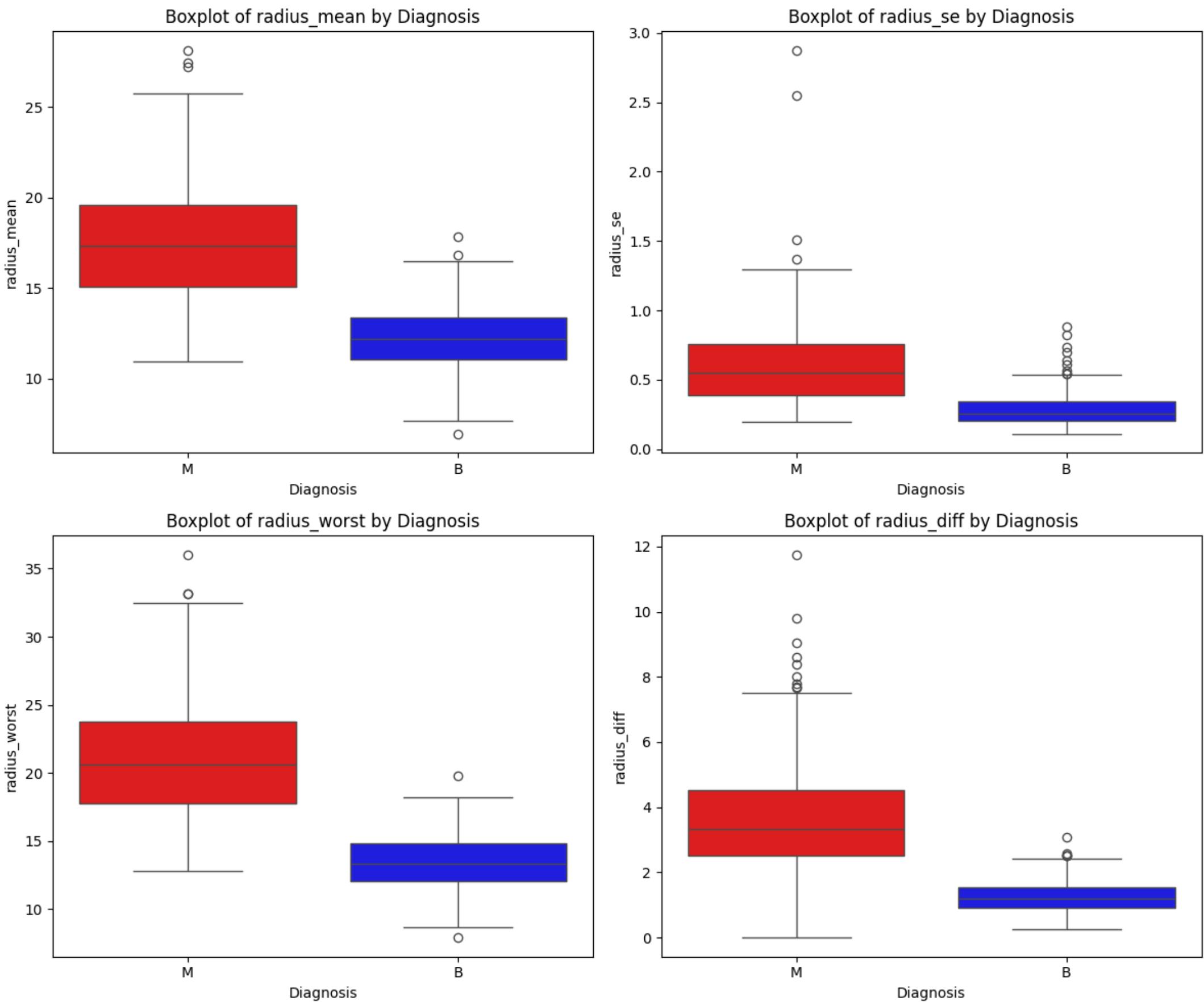
Overlaps

General observation:

Similarities in distribution of mean and worst, as well as of standart error and difference values for the radius

Why they matter?

Overlaps are the areas where it is questionable whether the patient should be diagnosed with benign or malignant.



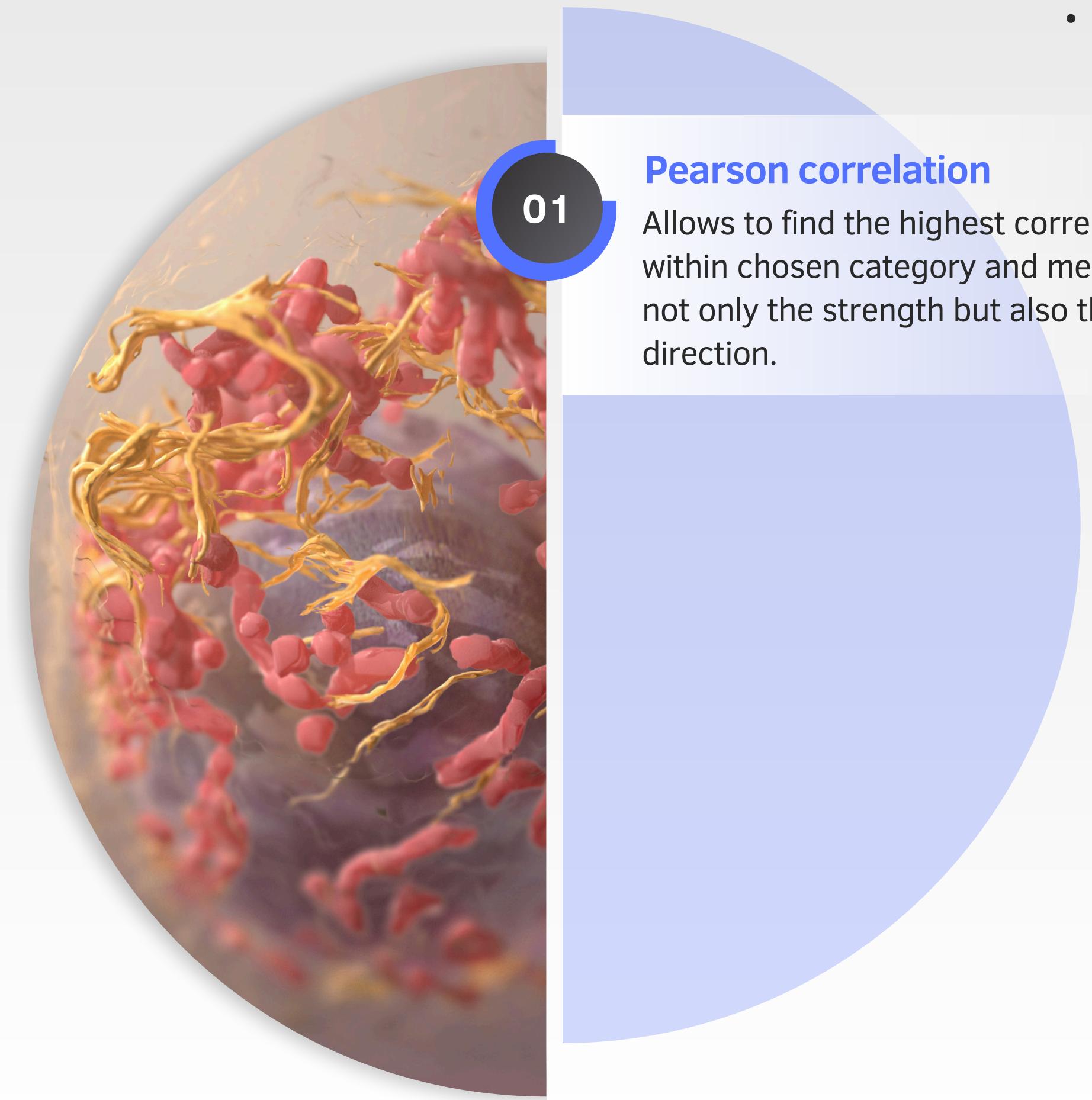


Top metric of the group

IDEA

Quantify diagnosis values and apply Pearson's to identify the highest correlations to diagnosis within the group.

	p-value	rPearson
radius_mean	8.465941e-96	0.730029
radius_se	9.738949e-50	0.567134
radius_worst	8.482292e-116	0.776454
radius_diff	6.444938e-89	0.711434





Top metric of the group

IDEA

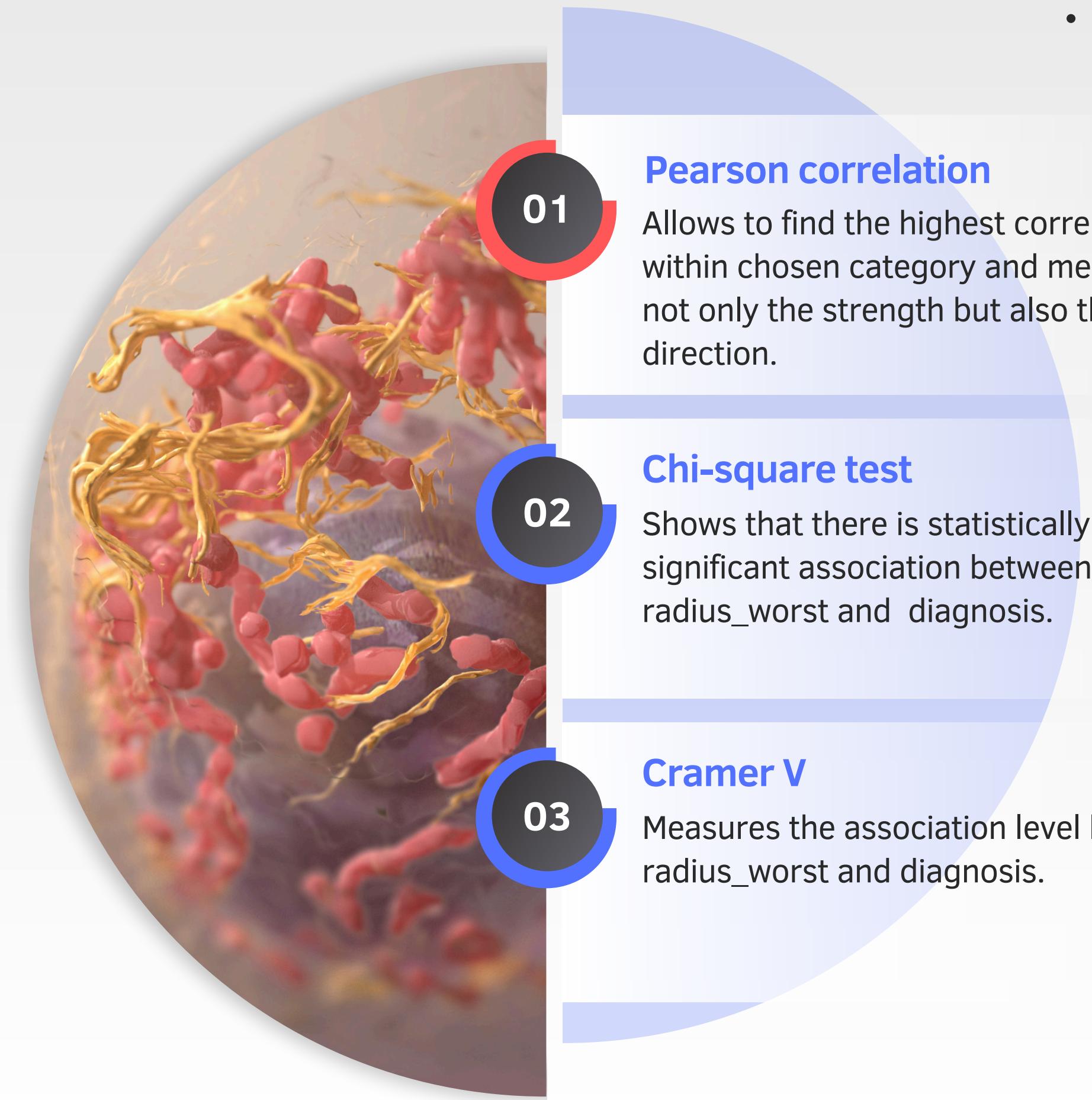
Discretize radius values using binning technique in order to apply Chi-square and Cramer's V tests.

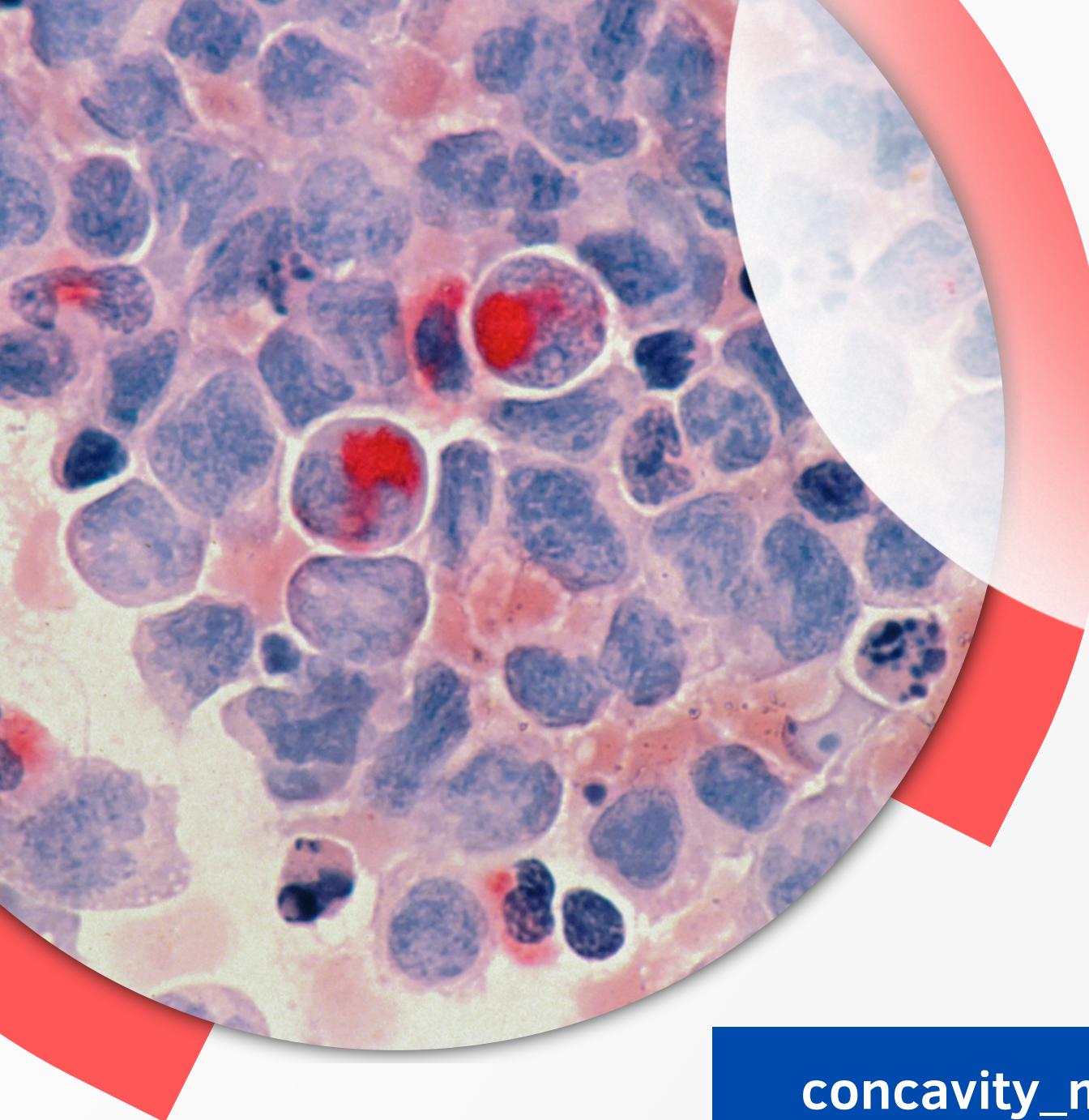
PURPOSE

Comparison with the other tops of their groups.

Chi-Square p-value is: 1.394378336723893e-26

Cramer V is: 0.679853961429309





⋮ ⋮ ⋮ ⋮ ⋮

Summary & Conclusions

concavity_mean	concave_points_worst	radius_worst	compactness_mean
0.74	0.71	0.68	0.58
2.022185081770995e-48	1.235542576364234e-39	1.394378336723893e-26	3.7694541006106594e-32

Thank you

