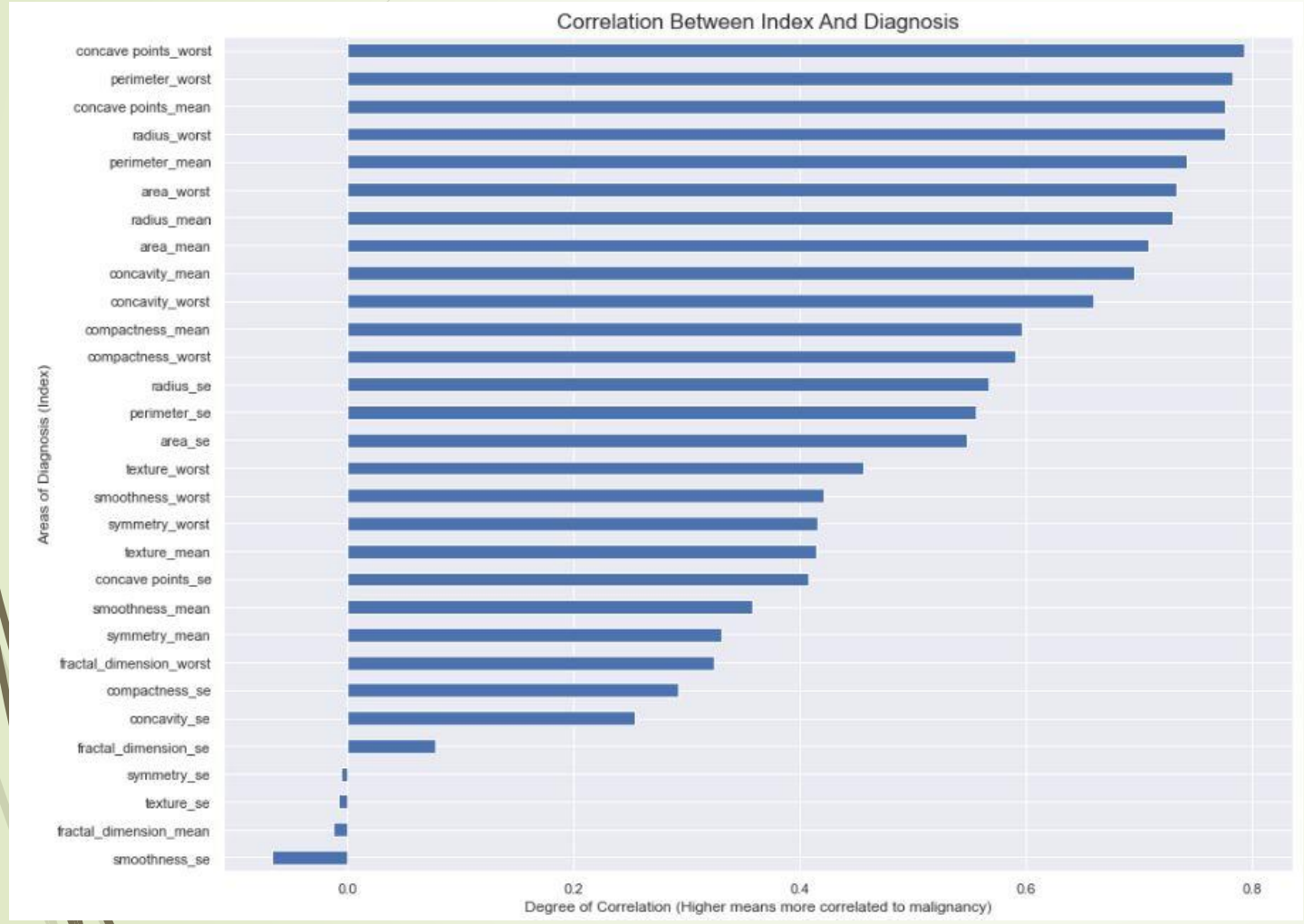
Breast Cancer Data & PREDICTION

https://www.kaggle.com/uciml/breast-cancer-wisconsin-data

Exploratory data analysis | Model training with Tensorflow & Keras API

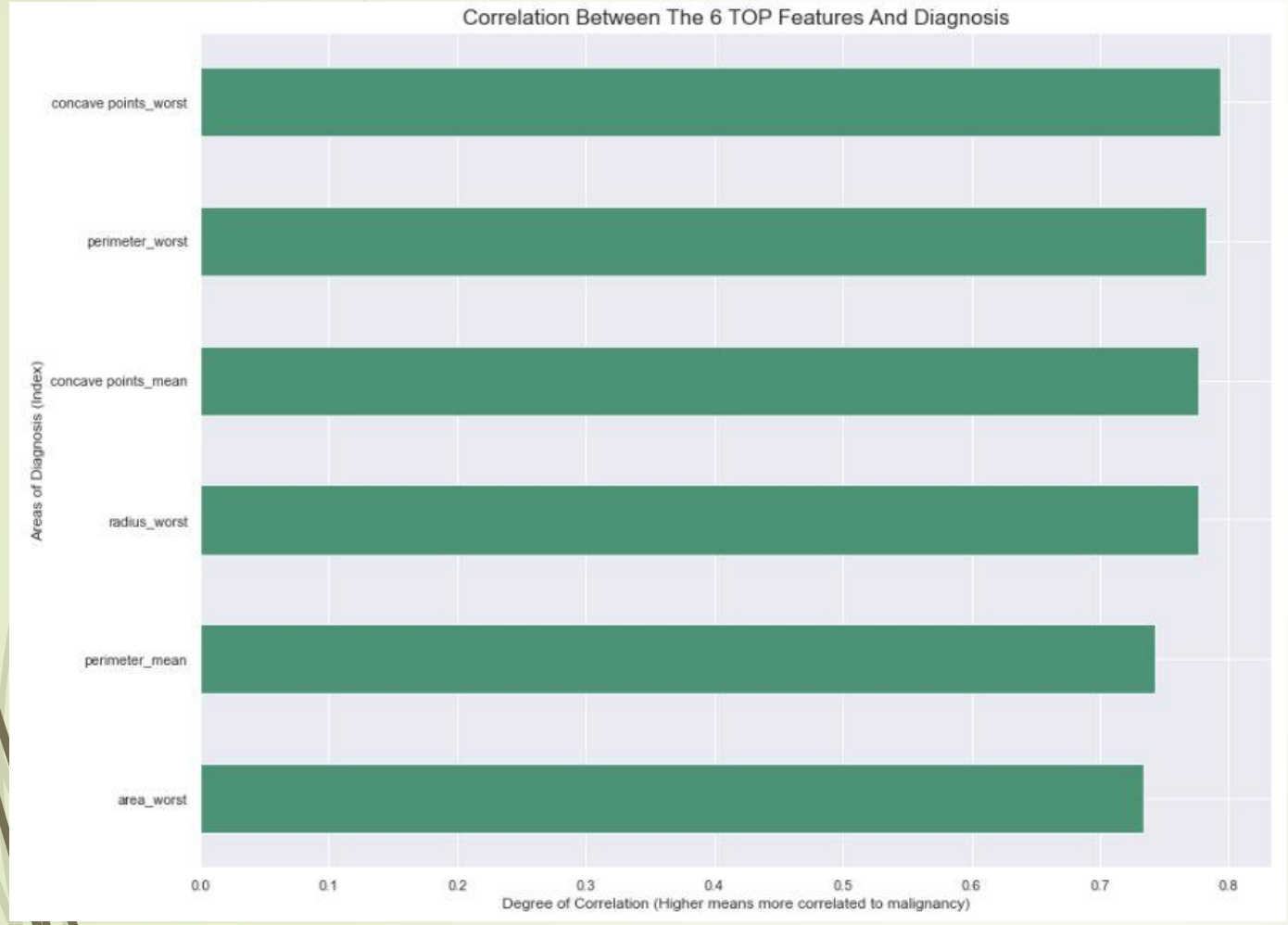
Do you know it is possible to predict whether your breast tumor is malignant or benign?



- Yes it is possible!
- There are 30 features or observations which are closely related to breast cancer.
- These are ranked with ‘Concave points_worst’ most related, and ‘fractal_dimension_se least related.
- And smoothness_se negatively related.

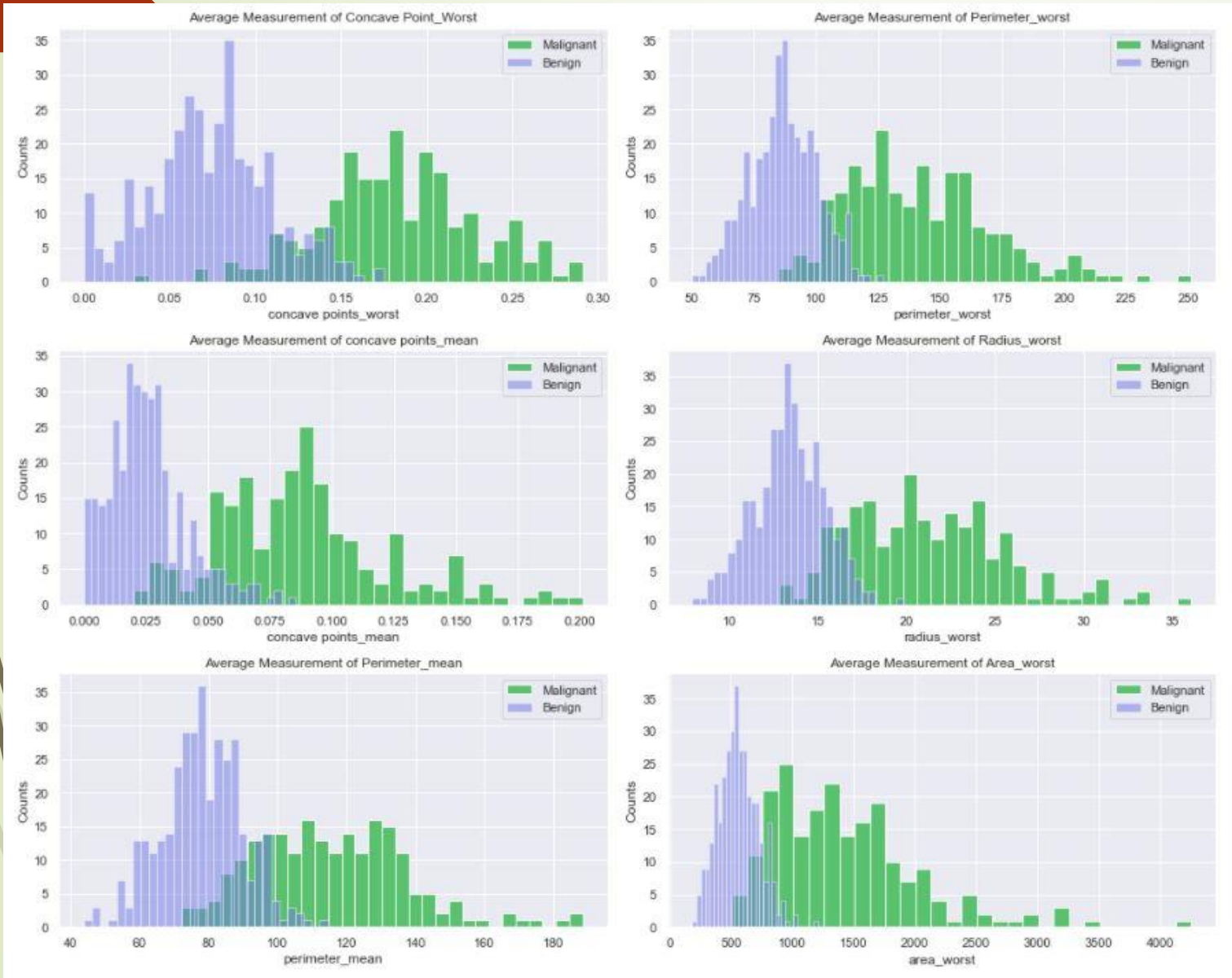
- Do I really have to go through all these measurements to determine my prognosis?

However there are TOP 6 features which you can consider before having to be tested on all features



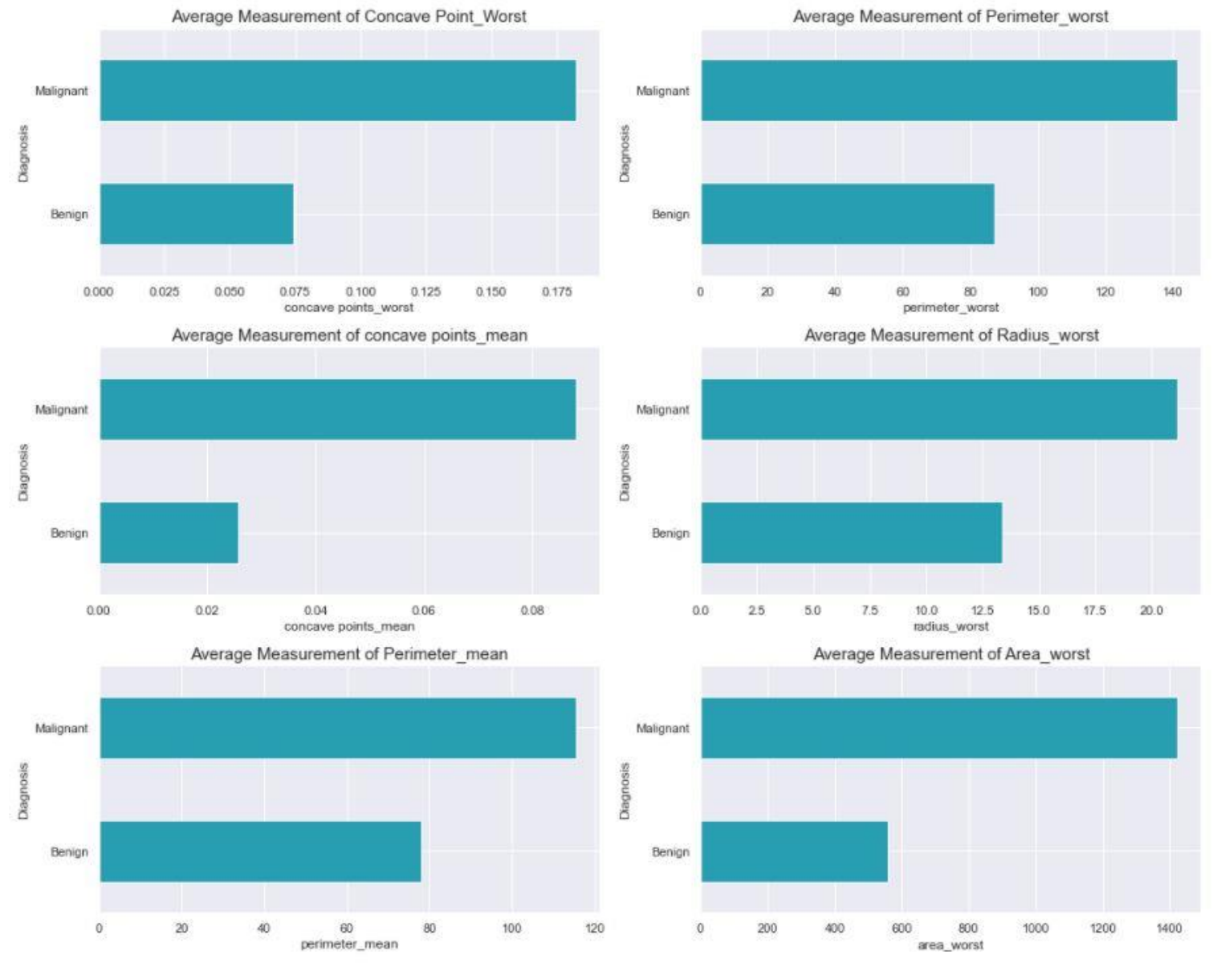
- Out of these 30 features, 6 have the strongest relationship with malignancy!
- In addition, 25 of these have measurements that defer significantly between malignant and benign.
- What do you mean?

The measurements of malignant tumor are significantly larger



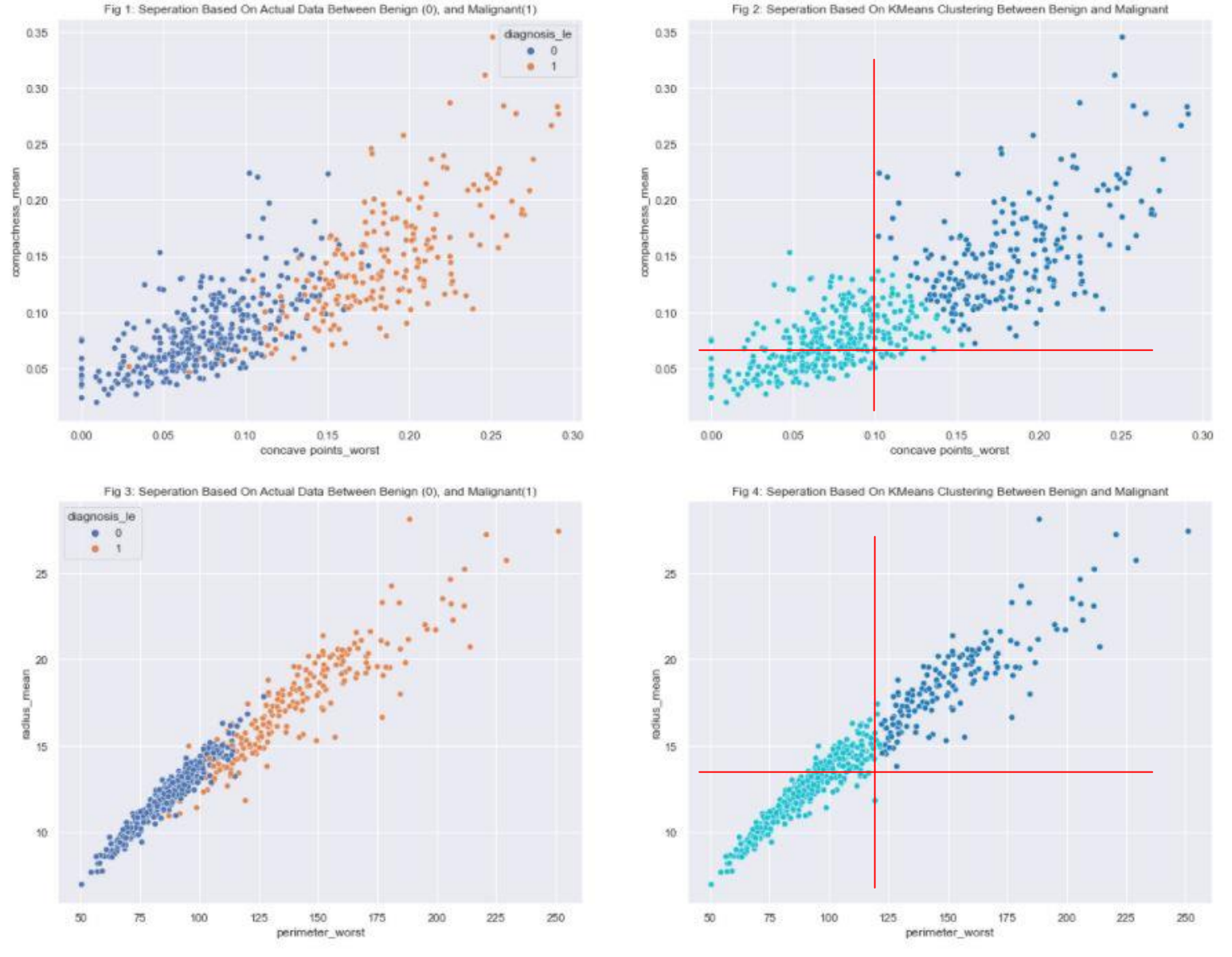
- The green graphs represent measurements of malignant tumor. The blue represents benign.
- You can see that the measurement defer significantly between the 2 groups.
- What do you mean?

The Average Measurements Are Also significantly Larger!



- To further illustrate what I mean, the graphs on the left show the average measurements of the tumor based on these 6 features.
- You can see that they defer significantly between the 2 groups… by at least 30%
- Is there a benchmark I should look out for?

We can make a prediction with these 6 features VS other features too!

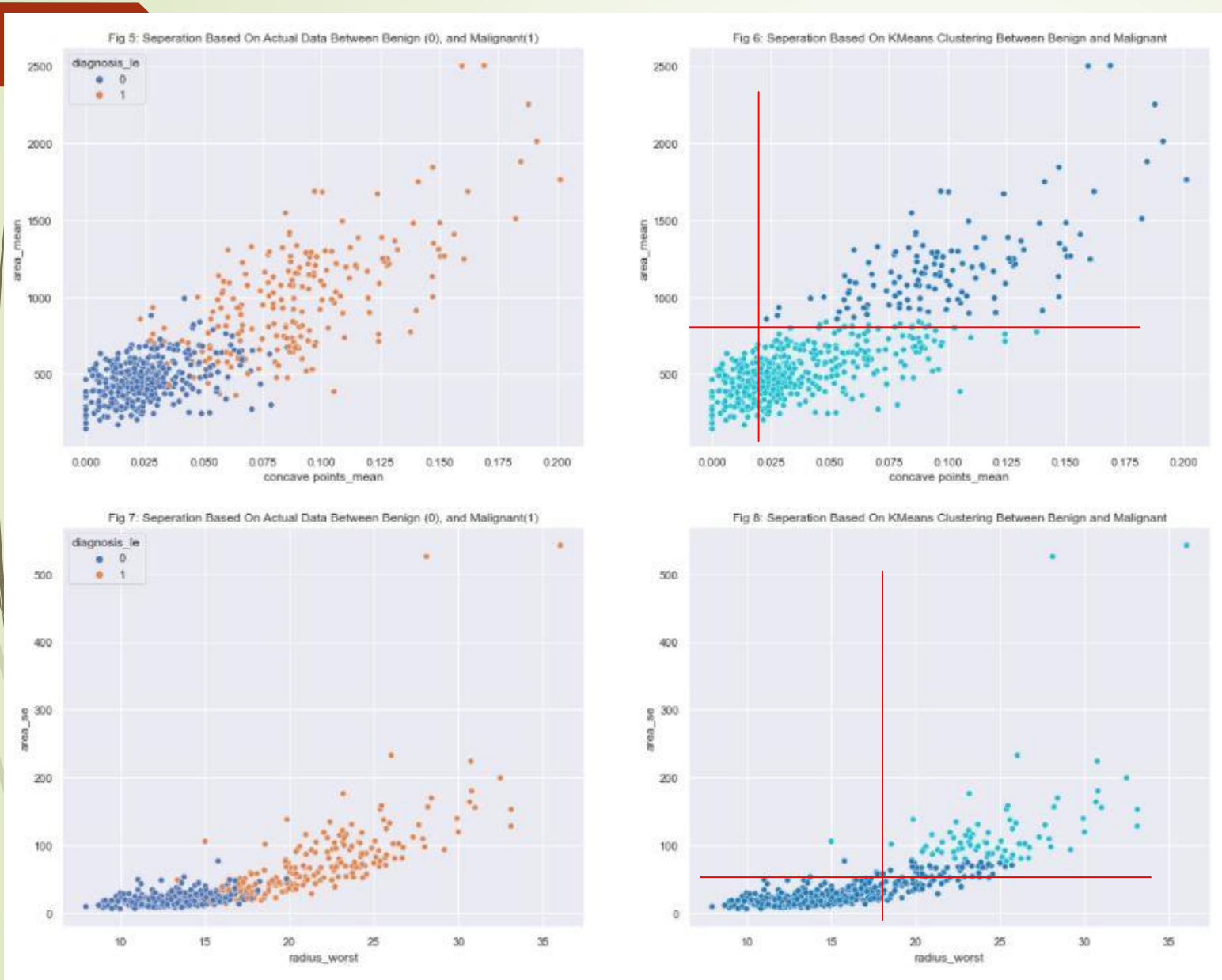


- Fig 1 and 3, shows the size of the tumor, and these represent actual historical results where the brown dots are malignant.
- Fig 2 and 4, shows the prediction of future tumors.

Tumor measurements (for benign):

- Concav points_worst < 0.10 and compactness_mean < 0.06. (Fig 2)
- Perimeter_worst < 120 and radius_mean < 14. (Fig 4)
- Are there anymore measurements?

Yes there are !



- Fig 5 and 7, shows the size of the tumor, and these represent actual historical results where the brown dots are malignant.

- Fig 6 and 8, shows the prediction of future tumors.

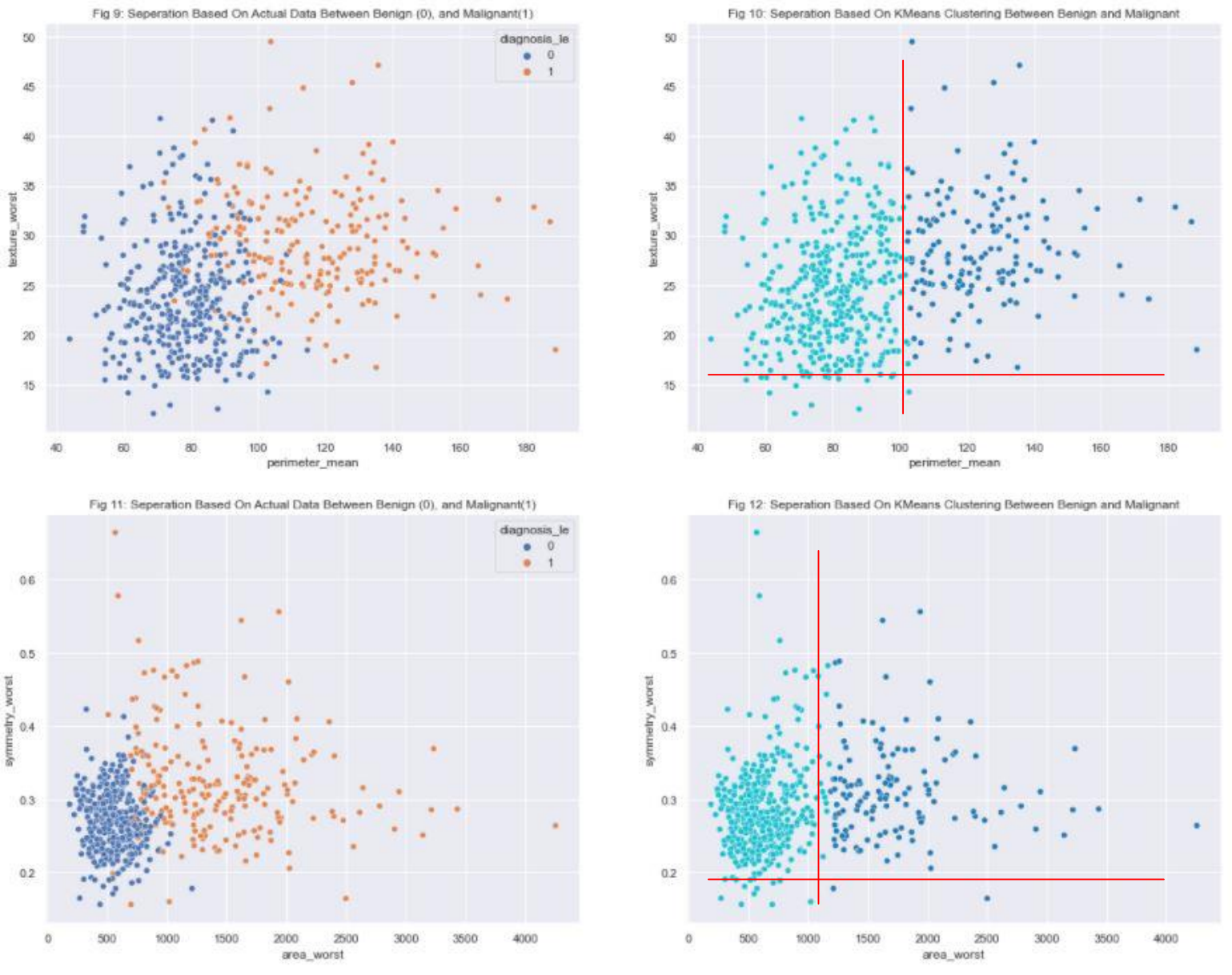
Tumor measurements (for benign):

- Concav points_mean < 0.02 and area_mean < 800 will be benign. (Fig 6)

- Radius_worst < 17 and area_se < 55 will be benign. (Fig 8)

- Are there anymore measurements?

Yes there are !



- Fig 9 and 11, shows the size of the tumor, and these represent actual historical results where the brown dots are malignant.

- Fig 10 and 12, shows the prediction of future tumors.

Tumor measurements (for benign):

- Perimeter_mean < 100 and texture_worst < 16 will be benign. (Fig 10)

- area_worst < 1100 and symmetry_worst < 0.15 will be benign. (Fig 12)

- Are there anymore measurements?

Summary & Conclusion

- There are 30 features or observations which are used to determine whether a breast tumor is malignant or benign.
- Out of these 30, 25 shows a significant difference in measurements comparing a malignant tumor with a benign one.
- The difference ranges between 30% -50% in their average measurements.
- By examining the 6 TOP observations we can accurately predict whether the tumor is malignant or benign.