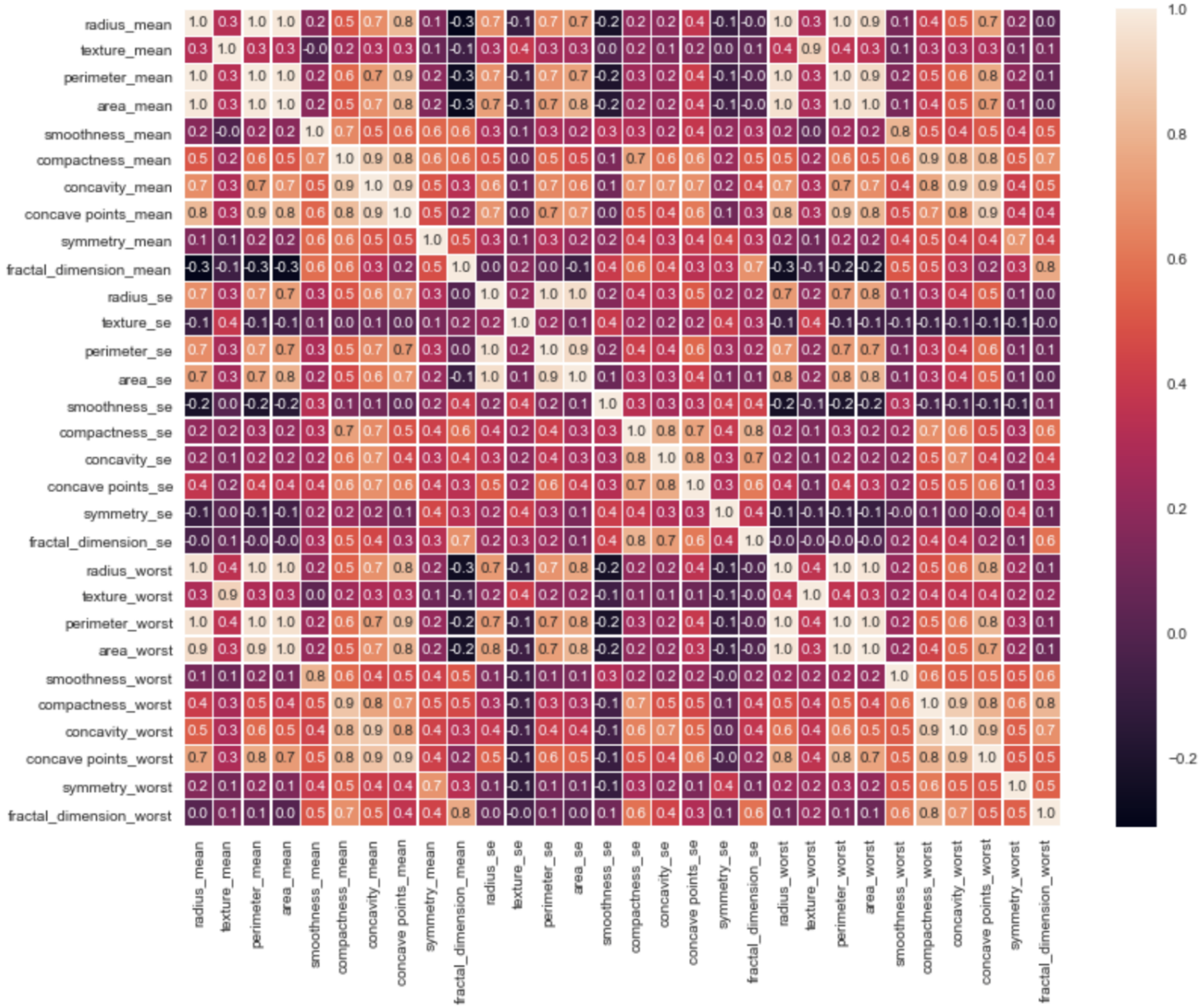
Supplementary Material

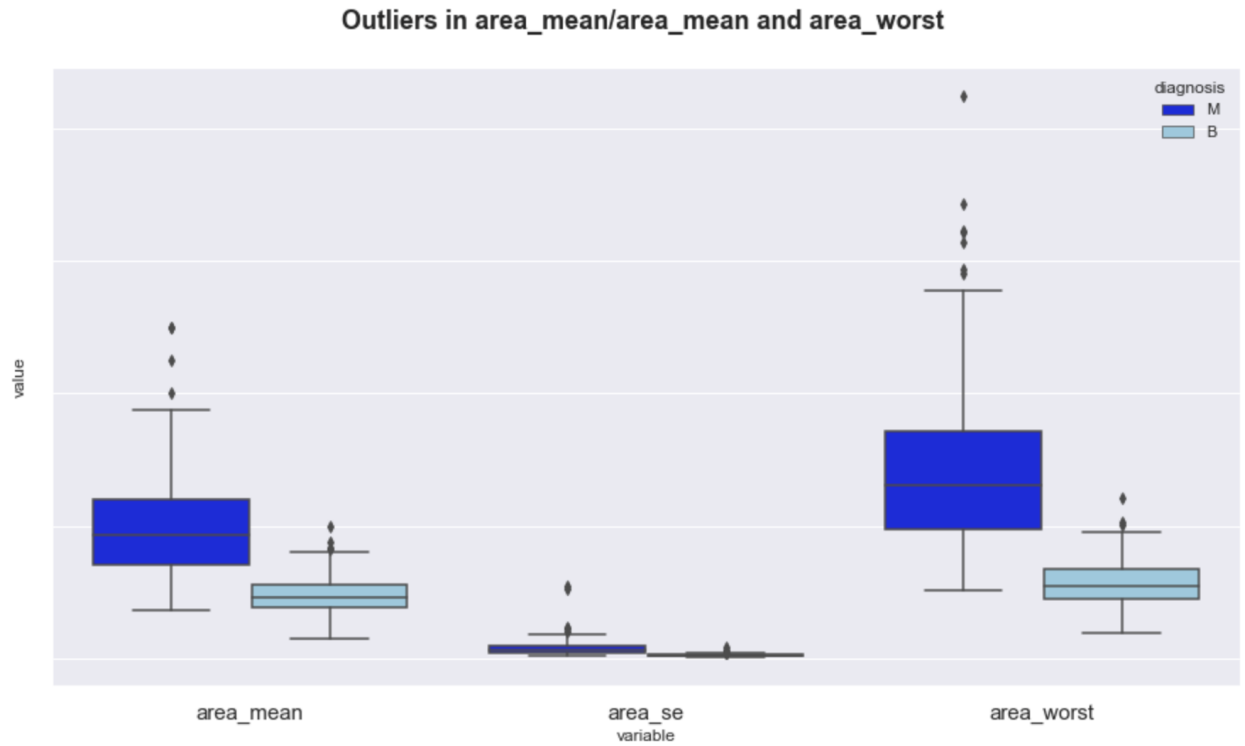


The figure 1 above shows the correlation matrix.

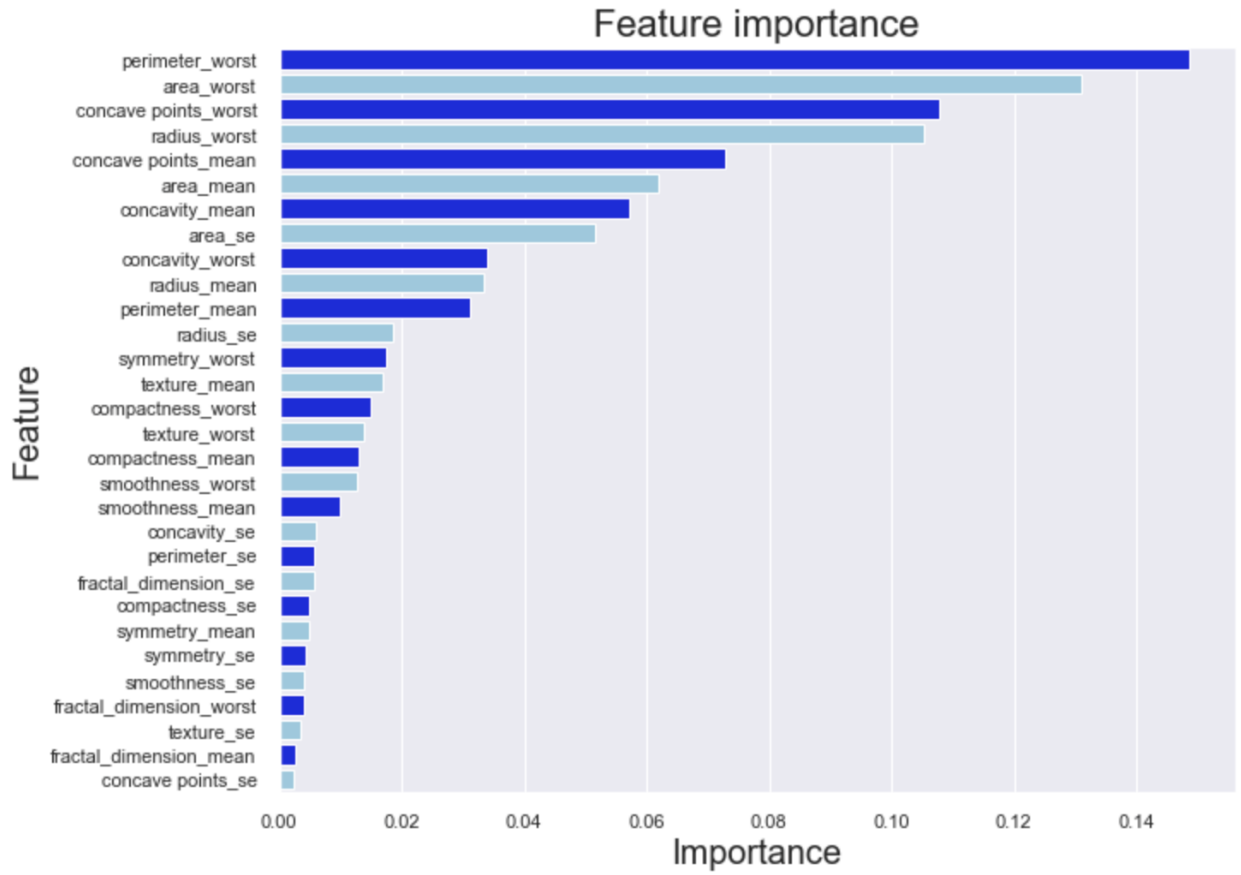
It can be observed from the above figure that the mean area of the nucleus has a strong positive correlation with the radius and parameter mean values. The correlation between some parameters (r between 0.5 and 0.7) is also moderately positive.

There are several types of data:

- Radius, measured from centre to point of perimeter (mean, standard error, worst[highest mean])
- Texture, measured by pixels making up the segmented area, is represented by grey-scale values (0 for black and 255 for white).
- Perimeter, which is the size of the core tumour
- Area, the area of the tumour
- Smoothness, measured by variation in radius length
- Compactness, measured by the mean of [(square of the perimeter)/area -1]
- Concavity, whether it is concave upwards or concave downwards
- Concave point (number of concave portions)
- Symmetry
- Fractal dimension, which measures complexity by comparing how a detail in a pattern changes with the scale at which it is measured.



This figure shows the outliers in 'area_mean' , 'area_se' and 'area_worst'



The figure above shows each feature importance in a graph form

Model prediction is affected by the importance of each feature (variable). The purpose of this measure is to determine how useful a specific variable is for a current model and prediction.