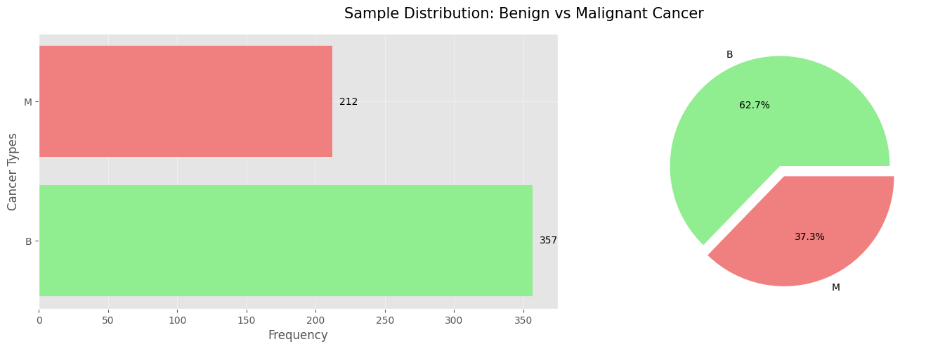
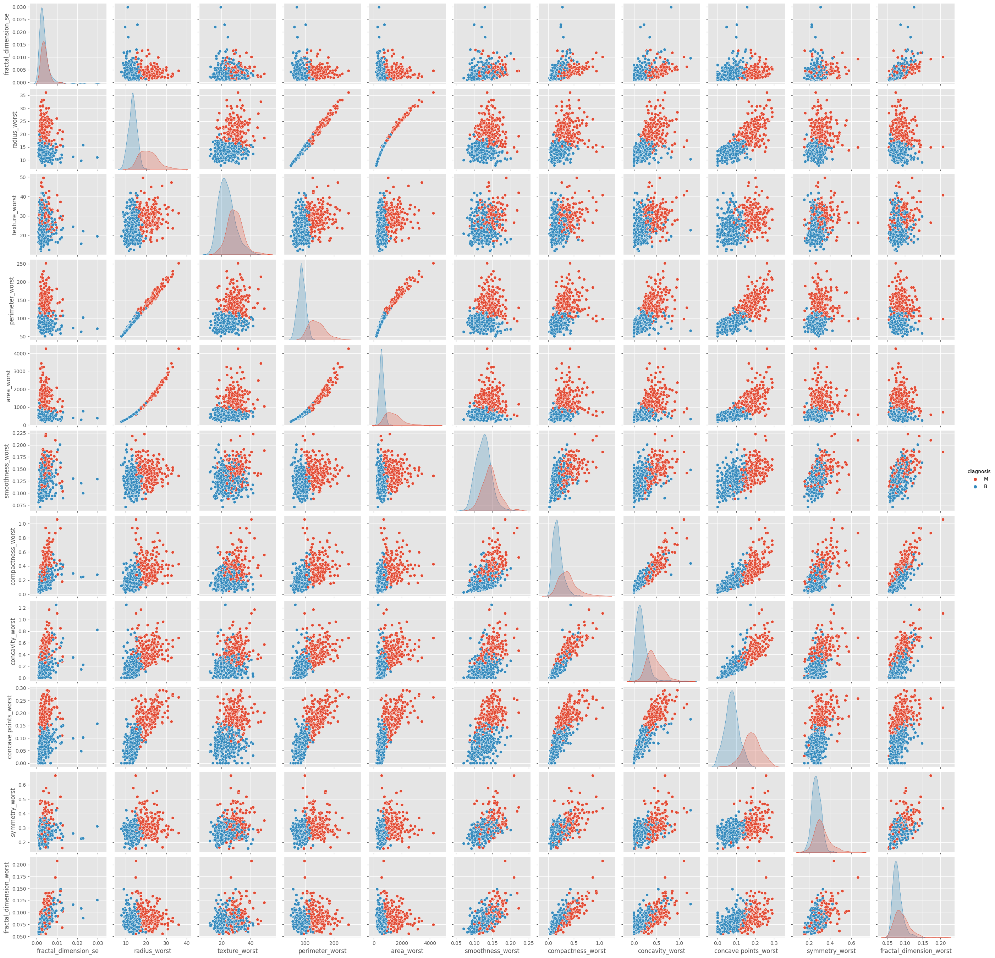
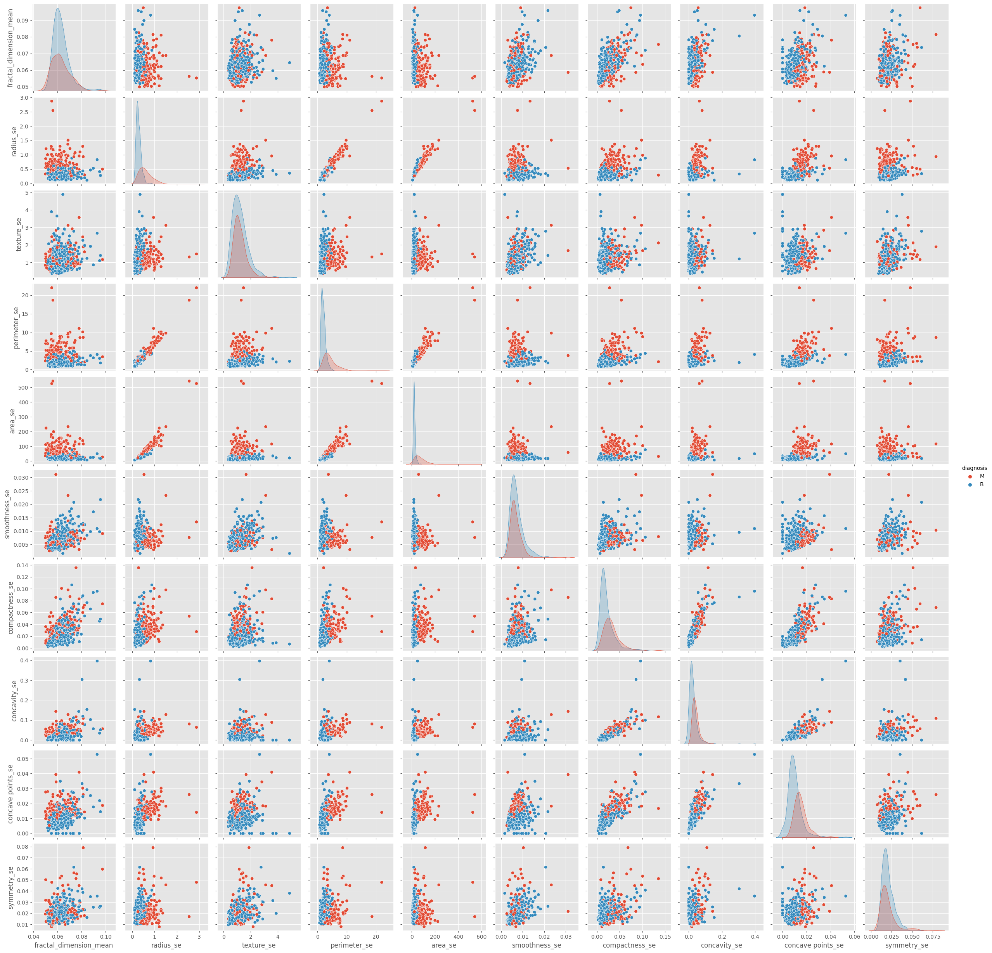
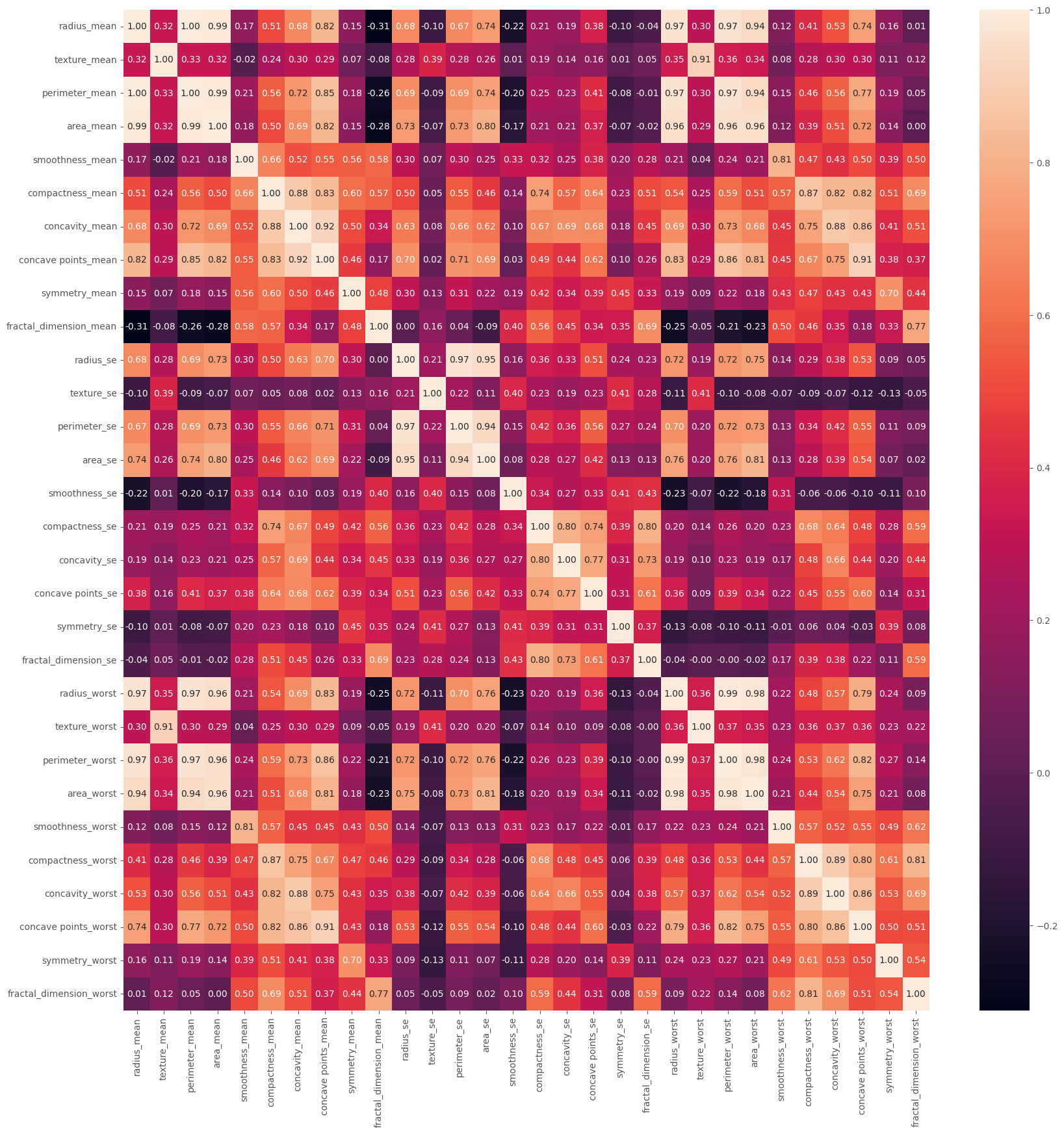
**Sample Distribution: Benign vs Malignant Cancer**



**Data Visualization**

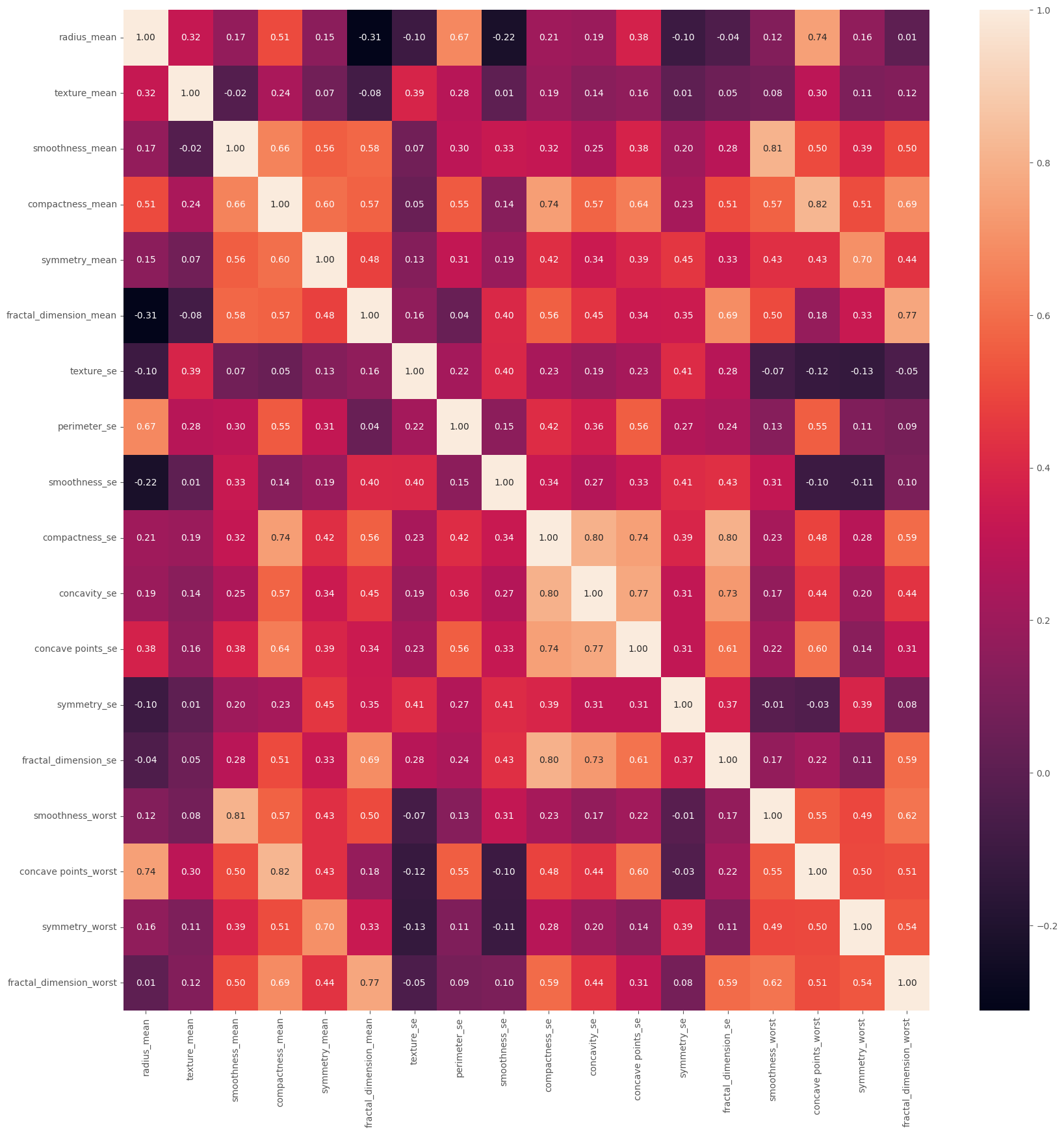


**# Result: radius\_mean - perimeter\_mean - area\_mean - concave points\_mean - radius\_worst - area\_worst - perimeter\_worst correlation is high,**

**# keeping radius\_mean is just enough.**

**# texture\_mean - texture\_worst correlation is high, keep texture\_mean.**

**# compactness\_mean - compactness\_worst - concavity\_mean, correlation is high. Keep compactness\_mean.**



**Model Training**

|  |  |
| --- | --- |
| **Model Accuracy, Precision, ROC\_AUC**  **Before Hyperparameter Tunning** | **Model Accuracy, Precision, ROC\_AUC**  **After Hyperparameter Tunning** |
|  |  |
| The following Model gives higher accuracy (desending order):   * Logistic Regression * LDA * SVC * XGB * Adaboost * KNN * RF * DT | For Voting Classifier used all classifiers as they are having higher accuracy and ROC\_AUC.  The output of Voting Classifier gives improved Accuracy, Precision and ROC-AUC value.  AUC of 0.99 falls into the "High discrimination" category. It indicates that the model has a high true positive rate and a relatively low false positive rate across different threshold settings, demonstrating its effectiveness in distinguishing between the positive and negative classes.  We are using Voting Classifier as it gives improved Accuracy, Precision and AUC value.     |  |  | | --- | --- | | Voting Soft All | Voting Hard All | |  |  | | True Positive: 104  False Positive: 4  True Negative:60  False Negative: 3  Precision = 0.96  96% of the correctly predicted case turned out to be positive.  Recall = 0.96  96% of the Positive were successfully predicted by the model. | **True Positive: 104**  **False Positive: 4**  **True Negative:59**  **False Negative: 4**  **Precision = 0.95**  **95% of the correctly predicted case turned out to be positive.**  **Recall = 0.95**  **95% of the Positive were successfully predicted by the model.** | |