# Classification Phase

Outliers, also known as anomalies, are data points that exhibit significant deviations from the rest of the dataset. The significance of outliers detection lies in its ability to mitigate the impact of inter-observer and intra-observer variability on data analysis and interpretation.

Inter-observer variability, stemming from differences in observations among different individuals, and intra-observer variability, reflecting variations in observations made by the same individual over time, can both introduce discrepancies in identifying outliers.

By employing robust outlier detection techniques, we can identify and address these discrepancies, ensuring that data points deviating significantly from the majority are appropriately accounted for. This process not only enhances the reliability and reproducibility of research findings but also reduces the influence of subjective interpretations on analysis results. We utilize various outlier detection methods, each tailored to identify anomalies within our dataset.

We utilized feature selection techniques to enhance the efficiency and interpretability of our models. One crucial aspect of feature selection is identifying collinear features, which are highly correlated with each other. Collinear features can introduce redundancy into the model, leading to overfitting and decreased generalization performance.

To address this issue, we employed a method to find collinear features based on the correlation coefficient between features. The correlation coefficient measures the strength and direction of a linear relationship between two variables, ranging from -1 to 1. A correlation coefficient close to 1 indicates a strong positive linear relationship, while a coefficient close to -1 indicates a strong negative linear relationship. A coefficient close to 0 suggests little to no linear relationship.
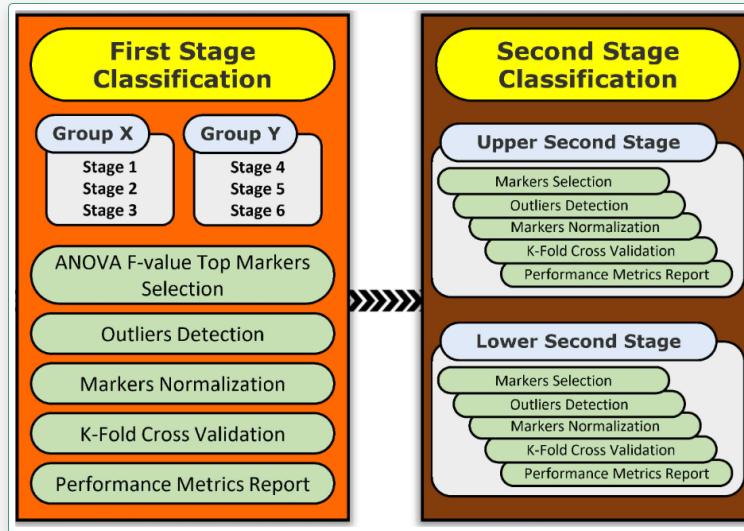
Current Extracted Features Files: 3
- ➜ Default_Dataset_All_Distances_Augmentation 2024-10-02 15:49:49
- ➜ Default_Dataset_All_Distances_No_Augmentation 2024-09-27 22:33:52
- ➜ ISBI_2025_All_Distances_Augmentation 2024-10-20 16:43:18

Extracted Features File:

Select Extracted Features File

Select Train-to-Test Ratio: (70% for training)

Apply Outliers Detection:

Yes

Outliers Fraction (10%)

Apply Features Selection:

Yes

Correlation Threshold (Features Selection) (90%)

Features Scaling (Checkboxes):
- ☑ No Scaling
- ☑ Standardization
- ☑ MinMax Scaling
- ☑ Robust Scaling
- ☑ Max Absolute Scaling

Classifiers (Checkboxes):
- ☑ Gradient Boosting Classifier
- ☑ eXtreme Gradient Boosting Classifier
- ☑ Light Gradient Boosting Machine Classifier
- ☑ Random Forest Classifier
- ☑ Decision Tree Classifier
- ☑ K-Nearest Neighbors Classifier
- ☑ Multi-Layer Perceptron Classifier
- ☑ Support Vector Classifier
- ☑ Logistic Regression Classifier
- ☑ Extra Trees Classifier
- ☑ AdaBoost Classifier
- ☑ Histogram Gradient Boosting Classifier

Number of Classification Trials to Produce %CI (Confidence Interval):

```
10
```

Perform Classification

Contact

```
10
```

Perform Classification

Contact