Random Forest (RF) will be used in this project since it is robust, interpretable, and has a strong performance on medical datasets. It operates based on the following mechanism:

**Component— Decision Tree**

Decision Tree is a binary tree structure using a recursive partitioning process to finish classification task. At each node, the data is split into two subsets with a feature and a threshold. The optimal split is determined by minimizing a node impurity measure: the Gini Impurity or Entropy.

Gini Impurity:

Where is the proportion of class samples in dataset. Value of 0 means pure (all samples in one class), and higher values means more mixed classes.

Entropy:

Entropy shows the uncertainty of a node; In a pure node entropy = 0.

The best split at each node is the one that maximizes the information gain:

Where is the impurity measure (Gini or entropy), and are the subset after splitting with the feature .

The splitting recursively is executed until the conditions are met (e.g., max depth)

**Selection of Random samples and Random features in RF**

Each tree is trained on a random subset of the data, which is drawn withreplacement from the original dataset. Mathematically:

This means each tree sees about 63% unique samples (the rest are duplicates).

In the subset of each tree, only a random subset feature is selected. This makes trees de-correlated, to avoid some strong features might dominate every tree.

**Aggregation and Prediction in RF**

After all trees are trained independently, their predictions are aggregated (only classification is discussed here):

Where B is the number of trees in the forest.

Because each tree is slightly different, averaging their predictions reduces random fluctuations and noise. As the number of trees B increases, the ensemble’s variance decreases, and the prediction stabilizes.

**Bias–Variance Tradeoff in RF**

A single decision tree tends to have low bias and high variance. In RF, variance is reduced by average many de-correlation trees. If is the average correlation between trees, the variance of the forest is approximately:

Where is the variance of an individual tree. As *B* (number of the trees) increase, the second term will vanish, and the variance only rely on .

The mechanism of RF shows that it is an ensemble learning algorithm combining multiple decision trees and is suitable for classification tasks. While individual decision trees may be overfitting in the data, a group of different trees, which are trained on different random samples and subsets of features, can achieve better performance.

In this project, RF is supposed to be effective because it can capture complex, nonlinear interactions in medical features (e.g., Age, Cholesterol, MaxHR, Oldpeak), while remaining robust to noise and outliers.